

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:11:48 ; Search time 16.6667 Seconds
(without alignments)
2781.424 Million cell updates/sec

Title: US-10-018-929C-3_COPY_650_1270
Perfect score: 3216
Sequence: 1 RAADAFILFGSSLNPSHDVHK.....SGNNKHFSGSSNISQNPADV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	162.5	5.1	1739	4	US-09-976-594-76
2	162.5	5.1	1739	4	US-09-538-092-824
3	162.5	5.1	1740	4	US-09-949-016-8860
4	147	4.6	1939	4	US-09-949-016-6925
5	147	4.6	1942	4	US-09-949-016-8135
6	146.5	4.6	630	4	US-09-248-796A-20275
7	145.5	4.5	1648	4	US-09-248-796A-14286
8	145	4.5	1960	4	US-09-538-092-1077
9	145	4.5	1960	4	US-09-949-016-10872
10	145	4.5	2733	4	US-09-949-016-11433
11	140.5	4.4	3210	4	US-09-538-092-1154
12	140.5	4.4	3248	1	US-08-353-700-1
13	140.5	4.4	3248	5	PCT-US95-16216-1
14	139.5	4.3	1427	4	US-09-538-092-1044
15	139.5	4.3	2482	1	US-08-328-254-6
16	138.5	4.3	1972	4	US-08-875-435B-4
17	138	4.3	1836	4	US-09-949-016-7432
18	138	4.3	3259	4	US-09-949-016-6507
19	135.5	4.2	8991	4	US-08-714-741-32
20	134	4.2	1388	4	US-09-976-594-296
21	134	4.2	1695	4	US-09-866-108A-15753
22	133	4.1	1388	2	US-08-685-576-1
23	133	4.1	1939	3	US-09-310-187A-1
24	133	4.1	1939	4	US-09-538-092-917
25	132.5	4.1	700	4	US-09-107-532A-5094
26	131	4.1	649	3	US-09-134-001C-3891
27	131	4.1	834	4	US-09-187-999-11

28	131	4.1	901	4	US-09-248-796A-14747	Sequence 14747, A
29	130.5	4.1	1401	4	US-09-750-590A-2	Sequence 2, Appli
30	130.5	4.1	1857	4	US-09-917-254-91	Sequence 91, Appli
31	130.5	4.1	1972	4	US-09-538-092-1084	Sequence 1084, Ap
32	130.5	4.1	1984	4	US-09-949-016-7111	Sequence 7111, Ap
33	130.5	4.1	1984	4	US-09-949-016-7112	Sequence 7112, Ap
34	130.5	4.1	1984	4	US-09-949-016-7113	Sequence 7113, Ap
35	130.5	4.1	2186	4	US-09-949-016-10828	Sequence 10828, A
36	130.5	4.1	2329	3	US-08-755-587-16	Sequence 16, Appl
37	130.5	4.1	2349	4	US-09-538-092-914	Sequence 914, App
38	130	4.0	1312	4	US-09-270-767-43669	Sequence 43669, A
39	129.5	4.0	1979	4	US-09-949-016-6468	Sequence 6468, Ap
40	129.5	4.0	2047	4	US-09-949-016-7404	Sequence 7404, Ap
41	128.5	4.0	963	4	US-09-914-259-22	Sequence 22, Appl
42	128	4.0	608	4	US-09-248-796A-19602	Sequence 19602, A
43	128	4.0	859	1	US-08-053-614-2	Sequence 2, Appli
44	128	4.0	859	1	US-08-316-397B-2	Sequence 2, Appli
45	128	4.0	859	2	US-09-034-306-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-976-594-76
; Sequence 76, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1
US-09-976-594-76

Query Match	5.1%	Score 162.5;	DB 4;	Length 1739;
Best Local Similarity	18.9%	Pred. No. 1.8e-05;		
Matches	152;	Conservative 116;	Mismatches 259;	Indels 279; Gaps 35;
Qy	2	ADAFILFGSSLNPSHDVHKIEKIESCSERTKIFRLYSVCTVEKA-----L	49	
Db	879	ADTVIFDSQNPQNDLQAQRAHRIGQKQVNIYRLVTKGTVEEIIERAKKXVLDHL	938	
Qy	50	ILAFQNNRQKAVEN-----LNRLTHALMAGASYLFDKLDHFHSSETPDGVSF	100	
Db	939	VIQMDTTGRTILENNGRNSNPNFKBELTALKFGAEDLFKELE-----GRES	988	
Qy	101	EQSIMGVIEHFSSILSSKGEENEVKLLLEAKHAQGTYSSTLFGEDHIKLSDEES	160	
Db	989	EPQEMD-----IDELRLAETRENEVSTATDELL-SQFKVANFATMEDEBEL-----BER	1038	
Qy	161	PNIFWSKLGCGNPMKYPSTDPORNRKVOYF-----GSEASPKTG	203	
Db	1039	PHKWDRII-----PEEQKVKVEERQKELEIYMLPRISSTKKAQTND	1085	
Qy	204	DGGNAKRRKASDDVTPRVTPDPPVDDERRKASGKHGALES PKVTLOSCKSSGTDG	263	
Db	1086	DSDTESKQQRSSASESET-----SDDKK-----PKRRGRPSVRKDLVEG	1130	
Qy	264	TLDGN-----DAFLYMGSHISGIPEDML--ASQNGKIPDESQRRLHTVLKPKMAKLC	316	

Db 1131 FTDAETRRFIKAYKCF-----GLPLERLECIARD-AELVDKSVADL-----KRLG 1174
QY 317 QVHLSDACTSMVGNFLEYVIEVNIHRIYEEBATTFOAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQVEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIHQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEVDYIYSILSCMSKSLFLEHTQGLQFCFGTNSK 422
Db 1231 MLHKSIPVDPEEKKKYCLTCRVKAHFDVWGVDDSRLLLLGIYEHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
Db 1290 -----LKLTDKILPVEVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGEAEKLLKR 1342
QY 465 -----CASHYSTATRDIEKTIISGIIKSKMSKSLFLEHTQGLQFCFGTNSK 501
Db 1343 KPRVKENKVPRLKEEHGIELSSPRHSDNPSGEVKKDDGLEKSPMKKKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KOKLETSKSVEA-----AVRITCSRTSTQVGD--- 539
Db 1402 EKQSSRKDKEGDKERKKSKDKKEPKSGDAKSSSKRSQGPVHITAGSEPVPIGEDED 1461
QY 540 -----LKLDDH-----NYERKFD-----BIKSEKNECLKS-- 564
Db 1462 DDLDDQETFSICKERMRPVKALQKDPKGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMHEVAKKLADEACWINRIKSWAAKLKVCV 598
Db 1522 DQSHIKLWRRNLWIFVSKFTFEDARKLHLYKWAHKRSQEE-----EQKKKDDV 1572
QY 599 PIQSGNNKHF-----SGSSN---ISON 617
Db 1573 ---TGGKKPFRPEASGSRDSLISQS 1595

RESULT 2

US-09-538-092-824
; Sequence 824, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqFormatter Version 0.9
; SEQ ID NO 824
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number O14647
US-09-538-092-824

Query Match 5.1%; Score 162.5; DB 4; Length 1739;
Best Local Similarity 18.9%; Pred. No. 1.8e-05;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
QY 2 ADALFGLSSLPNSHDVKKVEKIKIESCSRTKIFRLYVCTVVEKA-----L 49
Db 879 ADTVVFDSDWNFONDQAQRAHRICQKKQVNIYRLVTGTVVEEIEERAKKQWLDHL 938
QY 50 ILARQNMRONKAVEN-----LNRSLTHALLMWGASYLFDKLDHFHSSETPDGVSF 100
Db 939 VIQRMDDTTGRTILENNSGRNSNPNKEELTALKFGAEDLFKELE-----GEES 988

QY 101 EQSIMDGVIEHFSSILSSKGBENEVKLCILLLEAKHAQGYSSDSTLFGEDHIKLSDEES 160
Db 989 EPOEMD-----IDEIURLAETRENEVSTSATDELL--SQKVANFATMEDEBEL-----EER 1038
QY 161 PNIFWSKLLGKGNPMWKYPSDTPQRNRKRVQYPE-----GSEASPKTG 203
Db 1039 PHKDWDEII-----PEEQKKVEBEERQKELBEIYMLPRISSSTKKAQTND 1085
QY 204 DGNNAKRRKASDVTDPVTPVDDDDERKASGDHMGALSPKVITLOSCKSSGTDG 263
Db 1086 DSDTESKROAQRASASESETED---SDDDKK-----PKRGRPRSRVKDLVEG 1130
QY 264 TLQGN-----DAFGLYSMGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLPKMAKLC 316
Db 1131 FTDAETRRFIKAYKCF-----GLPLERLECIARD-AELVDKSVADL-----KRLG 1174
QY 317 QVHLSDACTSMVGNFLEYVIEVNIHRIYEEBATTFOAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQVEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIHQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEVDYIYSILSCMSKSLFLEHTQGLQFCFGTNSK 422
Db 1231 MLHKSIPVDPEEKKKYCLTCRVKAHFDVWGVDDSRLLLLGIYEHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
Db 1290 -----LKLTDKILPVEVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGEAEKLLKR 1342
QY 465 -----CASHYSTATRDIEKTIISGIIKSKMSKSLFLEHTQGLQFCFGTNSK 501
Db 1343 KPRVKENKVPRLKEEHGIELSSPRHSDNPSGEVKKDDGLEKSPMKKKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KOKLETSKSVEA-----AVRITCSRTSTQVGD--- 539
Db 1402 EKQSSRKDKEGDKERKKSKDKKEPKSGDAKSSSKRSQGPVHITAGSEPVPIGEDED 1461
QY 540 -----LKLDDH-----NYERKFD-----BIKSEKNECLKS-- 564
Db 1462 DDLDDQETFSICKERMRPVKALQKDPKGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMHEVAKKLADEACWINRIKSWAAKLKVCV 598
Db 1522 DQSHIKLWRRNLWIFVSKFTFEDARKLHLYKWAHKRSQEE-----EQKKKDDV 1572
QY 599 PIQSGNNKHF-----SGSSN---ISON 617
Db 1573 ---TGGKKPFRPEASGSRDSLISQS 1595

RESULT 3

US-09-949-016-8860
; Sequence 8860, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8860
; LENGTH: 1740
; TYPE: PRT
; ORGANISM: Human

US-09-949-016-8860

Query Match 5.1%; Score 162.5; DB 4; Length 1740;
Best Local Similarity 18.9%; Pred. No. 1.8e-05;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

Qy 2 ADAFILGSSINSHDVHVKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 880 ADTVIFDSDNPNQDLQAARHIGOKKQVNTYRLVTGKTVEEETIERAKKQWLDHL 939
Qy 50 ILARQNRQNKAVEN-----LNRSLTHALLMWSYLFDKLDHFSSETPDSGVSF 100
Db 940 VIQRMDTGTIRLENSCRSNPNFKEELTALIKFGAEDLFKELE-----GEES 989
Qy 101 EQSIMGVHFEFSSILSKGGEENEVKLCLLEAKHAQGYTSSDSTLFGEDHIKLSDEES 160
Db 990 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQFKVANPATMEDEBEL---EER 1039
Qy 161 PNIFWSKLLGKGNPMWKYPSDTPQNRKRVQYFE-----GSEASPTG 203
Db 1040 PHKQWDEII-----PEQRKKVBEERQKELEIYMLPRIRSTKKAQTND 1086
Qy 204 DGGNAKRRKASDDVTPRVDDDERKASGDHMGALSPKVITLQSSCKSSGTDG 263
Db 1087 DSDTESKRAQSSASETED-----SDDDK-----PARRGRPSVRKDLVEG 1131
Qy 264 TLDGN-----DAFLYSMGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLKPMKLC 316
Db 1132 FTDAEIRRFKAYKF-----GLPLERLCIARD-AELVDKSVADL-----KRLG 1175
Qy 317 QVHLSDACTSMWGNFLEYVIENHRIYEEPATTOAQFIALSWAALLVKOILSHKESLV 376
Db 1176 ELIH--NSCSAMQYEBQLEKASSEGKPGKR-RGPTIKISGV-QVNVKSIQHBEFE 1231
Qy 377 RANSEL-----AFKRSRVEDVIYSILSCMSLFLFHTQGLQFCFGTNSK 422
Db 1232 MLHKSIPVDPEKKYCYLTCRVKAHFDVEMGVDDSRLLGLIYERGV-NWELIKTDPE 1290
Qy 423 QSVVSTKLVBESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
Db 1291 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGEAKLKR 1343
Qy 465 -----CSHYSTATRDIKTIGIKKKYKQVQKLVQEHBEKK 501
Db 1344 KPRVKKENKVPRLKEHGIELSSPRHSDNPSEGEVDDGLEKSPMKKQK-KKENENK 1402
Qy 502 MELLNMYADK-----KQKLETSKVEA-----AVIRITCSRTSTQVGD--- 539
Db 1403 EKQSSRKDEGDKERKSKDKKPKSGDASKSSKRSQGPVHITAGSEPVPIGSEDD 1452
Qy 540 -----LKLIDH-----NYERKFD-----BIKSKNECLKS-- 564
Db 1463 DDLDOETFSICKERMPVKALKQLDRPKGLNVQEQLEHTRNCLLKGIDRIAECLKAYS 1522
Qy 565 -----LEQMEHVAKKLADEACWINRIKSWAAKLKVCV 598
Db 1523 DQEHIKLWRRLMTFVSKFTFEDARKLHLYKMAHKRSQEBE-----EQKKKDDV 1573
Qy 599 PIQSGNNKHF-----SGSN-----ISON 617
Db 1574 ---TGGKKPFRPEASGSSRSLISQS 1596

RESULT 4

US-09-949-016-6925
; Sequence 6925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASES, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949.016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6925
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6925

Query Match 4.6%; Score 147; DB 4; Length 1939;
Best Local Similarity 20.3%; Pred. No. 0.00063;
Matches 137; Conservative 104; Mismatches 257; Indels 178; Gaps 27;

Qy 19 KHVEKIKIESCSER-TKIFRLYSVCTVEEKALILARQNRQNKAVENLNRSLTHALLMWG 77
Db 972 KHATENKVKNLTEEMAGLDETIAKLTKKKALQEAHQOTLDDLMQEE----- 1018
Qy 78 ASYLPDKLDHFSSETPDSGVSEQSIWDGVIHEFSSILSKGGEENEVKLCLLE--AK 135
Db 1019 -----DKVNTLTAKT-----KLEQQVDD-----LEGSLEQKKLCMDLERAKR 1057
Qy 136 HAQO--TYSSDSTLFGS-DHIKLSDESPNIFWSKLLGKGNPMWKYPSDTPQNRKRVQ- 191
Db 1058 KLEGDLKLAQESTMTDENDKQNLNEKLKKKEFENSLQKTEDEQALAMOLQKKIKELQA 1117
Qy 192 YFEGSEASPTGGGNAKAKKASD-----DVTD-----PRVTDPPVD--DDEKASGD 239
Db 1118 RIBELEIEAERASRAKQKRSDLRSELEISERLEEAGGATSAQITLNNKKREAEFOK 1177
Qy 240 HMGALSPKVITLQSSCKSSGTDGTLGNDAFLYSMGSHISGIP----- 284
Db 1178 MRRDLE-----ESTLQHEATAAALKKKHADSVAELGKQIDSLQRVKQKLEKSKSELKM 1230
Qy 285 -----EDMLASQDWGKIPDESQRRLHTVLKPMKLC--- 315
Db 1231 EINDLASNMETVSKAKANPEKMCETLEDQISEI---KTKEEQQLINELSAQKARLHTE 1287
Qy 316 -----CQVHLSDACTSMWGN-----FLEYVIENHRIYEE-----PATTFOAFOIALSWIAAL 363
Db 1288 SGFSPRLDEKDAVMVSQSLRGKQAFQOI1BELKRLQEEETKAKSTLAHALQSA-RHDCDL 1346
Qy 364 LVKQILSHKES-----LVYRANSELAFKCSRVEDVIYSILSCMSLFLFHTQGLQFDC 416
Db 1347 LREQYEEQAKAELQGMKANSKSEVAQWRTKYETDAI-----QRTBELE--- 1391
Qy 417 FGTSKQSVVSTKLVBESLSGATVRDEKINTK-----SMRNSSEDEBCH---TEKR 464
Db 1392 -----EAKKLAORLODAEHEVAVNSKCSASLEKTKQRLQNEVEDLMIDVRSNAA 1442
Qy 465 CSHYSTATRDIKTIGIKKKYKQVQKLVQEHBEKK-----MELLNMYADKKQKLETS 518
Db 1443 CIALDKKQRFDKVLAEWKQKYEETQAELEASQESRSLSTELFKVKNAYEESLDHLETL 1502
Qy 519 KSVFAAVIRITCSRTSTQVGDKLLLDHNYERKPEDEIKSEKNECLKSLEOM-----HEVAK 573
Db 1503 KR-ENKNLOQEI1DLTEQIAEGGKHIIHELEKVKQKQDHEKSELSQTSLEEAASLEHEEGK 1561
Qy 574 KCLAEDEACWINRIKS 589
Db 1562 ILRIQLE---LNQVKS 1574

RESULT 5

US-09-949-016-8135
; Sequence 8135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8135
; LENGTH: 1942
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8135

Query Match 4.6%; Score 147; DB 4; Length 1942;
Best Local Similarity 20.0%; Pred. No. 0.00063;
Matches 134; Conservative 101; Mismatches 270; Indels 164; Gaps 24;
Qy 19 KHVEKIKIESCSER-TKIFRLYSVCTVEEKALILARQNMQRQKAVENLNRLSHALLMVG 77
Db 975 KHATENVKVNLTEBEMAGLDETIAKTKEKALQEAHQOITLDDLOME-----1021
Qy 78 ASYLFDKLDHFHSETPDGVSFQSGIMDGVHFFSILSSKGGEEVVKCLLLE--AK 135
Db 1022 -----DRVNTLTAKT-----KLEQVDD-----LEGSLEQKKLCMDLERAKR 1060
Qy 136 HAQG--TYSSDSTLFG- DHIKLSDESPNIFWKLGGKPMWKPYSPTPQNRKRVQ- 191
Db 1061 KJEGDLKLAQESTMDTENDKQNLKXKKEFEFMSNLQKTEDEQALAIQLKXKELQA 1120
Qy 192 YFEGSEASPTGDCGNKAKKASDDVTPRVTDPVDDDERKASGDHMGALSPKVIT 251
Db 1121 RIELEIEIEAERASRAKQKQSDLSRELEIEISERLEEAGGATSAQIEMWKKREAFQK 1180
Qy 252 L-----QSSCKSGTDTGNDAGLYSMGSHITGIP-----284
Db 1181 MRDLEESTLQHEATAAARAKKHADSVAEELGEQIDSLQKVQKLEKEKSELKMEINDLAS 1240
Qy 285 -----EDMLASQDWGKIPDESORRLHTVLKPKMAKL-----CQV 318
Db 1241 NMETVSKAKANFERKMCRTLEDQISEI---KTKEBEOQLNELSAQKARLHTSGEFSRQ 1297
Qy 319 LHLSDACTSMVGN---FLEYVIENHRIYEE---PATTFOAFQIALSWIAALIVKQILS 370
Db 1298 LDEKDMVSLSRGKQAFQTOIEELKQLEBEETKAKSTLAHALQSA-RHDCDLLREQYEE 1356
Qy 371 HKES-----LVANSSELAPKSRVVDVYIYSILSCMSLFLBHTQGLQFDCFGTNSKQ 423
Db 1357 EQEAKAELQRMGSKANSEVAQWRTKYETDAI-----QRTBELE-----1394
Qy 424 SVVSTKLVSLSGATVRDEKINTK-----SMRNSSEDEECM-----TEKRCSHYSTA 471
Db 1395 --EAKQKLAORLQDAEBHEAVNSKCLASLEKTQRLQNEVEDLMDIVERSNAACIALDKK 1452
Qy 472 TRDIKTIISGKKYKQVOKVLQVEHEBK-----MELNMVADKKQKLETSKSVEAAV 525
Db 1453 QRFDFKVLAEWKQYETQAELEASQKESRSLSTELFKVKNAYEESLDHLETAKR-ENKN 1511
Qy 526 IRTICSTSTQVGLKLLDHNHYERKFDIEIKERKNELCKLSQOM-----HEVAKKGLAEDE 580
Db 1512 LQOEISDLTEQIABGGKHIELEKVKVQLDHEKSELQTSLEAEASLEHBEGKILRIQLE 1571
Qy 581 ACWINRIKS 589
Db 1572 ---LNQVKS 1577

RESULT 6

US-09-248-796A-20275
; Sequence 20275; Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20275
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20275

Query Match 4.6%; Score 146.5; DB 4; Length 630;
Best Local Similarity 20.5%; Pred. No. 0.00011;
Matches 125; Conservative 89; Mismatches 210; Indels 187; Gaps 26;
Qy 108 VIHFPSILSSKGGEEVVKCLLLEAKHAQ--GTYSDDSTLFGEDHDKLSDESPNIFW 165
Db 40 VIKEYSDKIKLSKINSIK-----ENHSKEITTHNEOKTSLKQDIKLSQDH--ELAQ 91
Qy 166 SKLLGKGNPMWKPYSPTPQNRKRVQYFEGSEASPTGDCGNKAKKASDDVTPRVT 225
Db 92 TQLEDKENQLKELKASLEKHN-----TESATSIEEKN--QIKELSETIKSLK-TE 139
Qy 226 PVDDDERKASGDH-----MGALSPKV-----ITL 252
Db 140 LKTSGDALKQSQKQYKTLTKNSDTELKLEKQLEELKVKLDLQTADEKLKGITERIAL 199
Qy 253 QS---SCKSGTDTG-----LDGNDAFLGYSMGSHISGIPEDM 287
Db 200 KSELETVMKNSGLSTTSELAAALTKVKSLEKEKEELOFLSGNKSKELEDYIQKHSIDISEKL 259
Qy 288 LASQDWGKIP---DESORRLHTVLKPKMAKLQVLHLSDACTSMVGNFLEYVIENHRIY 343
Db 260 KALTDELKKTQFDDSKKKL-TELENDLTSTKKELETEKTQTSKPKN-----LEERKDK 313
Qy 344 EEPATTFOAFQIALSWIAALIVKQILSHKESLVANSSELAPKSRV--EVDYIYSILSCM 401
Db 314 E-----IVKLNKELELLKNDNSGAKKELLEKVKLSLESEIEILSKSELEDK 357
Qy 402 KSLFLSH-----TQGLQFD-----CFGTNSKQSVWSTKL 431
Db 358 KSVMKQHDELKEQTKENKQELQKVTDYSTTKLKLDELQKELDAALSFKDKPETAACL- 416
Qy 432 NESLSGATVRDEKINTKSMRNSSEDEECMTEKRCSHYSTATR-----DIETIGIGIK 484
Db 417 TQSTDSLEAAKKNLI--LISEKEKTEQELKUTKQHALEKLLDDKNADSSKNTSDLEK 474
Qy 485 KYKK-----QVOKLVOEHEEKQKELLNMYADKQKLETSKSVEAAVIRTCSTSTQVG 538
Db 475 ANKVLDDLEQLEKATKELDTNKL------QKHEKL-----NNEFT 511
Qy 539 DLKLLDHNHYERKFDIEIKERKNELCKLSQOM---HEVAKKGL--AEDE-----ACWINRIK 588
Db 512 NLSKSHHEIKTKLDKLEKNEKNLDLVKSHKELNVEHKEKSTKLVNVEBENKKITAKYSELK 571
Qy 589 SWAAKLKVCVP 599
Db 572 KLQDKLKNVCVP 582

RESULT 7

US-09-248-796A-14286


```
Db 1504 TEMEDLSSKDDVGKS---VHELEKSKRALEEQ--VEEMKTQLEEELEDELOATEDAKLRL 1558
Qy 597 CVPIQS 602
Db 1559 EVNLQA 1564

RESULT 9
US-09-949-016-10872
; Sequence 10872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10872
; LENGTH: 1960
; TYPE: PRN
; ORGANISM: Human
US-09-949-016-10872

Query Match 4.5%; Score 145; DB 4; Length 1960;
Best Local Similarity 18.3%; Pred. No. 0.00099;
Matches 133; Conservative 107; Mismatches 220; Indels 266; Gaps 25;

Qy 22 EKIKIESCSERTKIFRLYSVCTVEKALILARONRQNKAVENLRSLTHALLMWGASYL 81
Db 960 QKQLEKVTTEAKKKU-----EEQIILEDQCKLAK----- 992
Qy 82 FDKLDHPSSETPDGVSFQSGIMDGVHFEFSSILSSKGGEEVVKLCLEAKHAQGTY 141
Db 993 -----EKKLEDRIAEFTNLTE-----EEKSKSLAKLKNKH----- 1025
Qy 142 SSDSTLFGEHDHILKLSDESPNIFWSKLLGGKQPMWKYPSDTPQNRKRQVYFEGSEASPK 201
Db 1026 -----EAMITDLE-----ERLRREEKQROELE----- 1047
Qy 202 TGDGNAKRRKASDDVDTPRVTPPPVDDDERKASGKDHMGALSPKVTILQSSCKSSGT 261
Db 1048 -----KTRKLEGDSTD--LSQIAELQAIKELQWLAKKE-----ELQAAARVEE 1094
Qy 262 DGTLDGNDAFGLYSMGSHISGIPEDMLASQDWGKIPDESQRRLHTVLKPKWAKLCQVLHL 321
Db 1095 EAAQKQWALKKIRELSQISELQELSDLESASRNKAQKQKDLGEELEALKTELEDTLDS 1154
Qy 322 SDACTSMVGNFLEVIENRIYEPATTFQAFQI-----ALSNIAALL----- 364
Db 1155 TAAQQLRSKREQEWNILKTKLEBEAKTHEA-QIQEMRQKHSQAVELEAQLQETKRVKA 1213
Qy 365 -----VKOIL-----SHKESLV-----RANSELAF 384
Db 1214 NLEKAKQTLNERGELANVKVLQGGDSEHKKKVBAQLQELQVFNQGERVETELAD 1273
Qy 385 KCSR--VEVDYIYSILSCMSLFF-----LEHTQGLQDFCGTNSKQSVVSTK 429
Db 1274 KVTQLQVELDNVTVLLSQSDSKSLTKDFSALESQLDQTQLLQEQ---ENRQKLSLSTK 1330
Qy 430 LVNESLSGATVRDKINTKSNRNSSEDEECMTEKRCCHYSTATPDIEK-----TI 479
Db 1331 L-----KQVEDEKNSFREQLEEBEAKHNLKQIATLHAQVADMKKQMEDSVGCLETA 1383
Qy 480 SGIKKYKKQVQKLQVEHEEK-----KME-----LLNM----- 507
```

```
Db 1384 EEVKKLQKQDLEGLSQSHEEKVAAVYDKLEKTKTRLQOELDDLVLVDLHQRQSACNLEKKQ 1443
Qy 508 -----YADKKQKLET---SKSVEAAVIRITCSRTSTQVGDKLALDHNYE 548
Db 1444 KKFQDLAEEKTISAKYAEERDRAEAAREKTKALSARALEEAMEQKAEERLNRKQFR 1503
Qy 549 RKFDEIKSEKNECLKSLEQMHVAKKKLADEACMINRIKSWA-----AKLKV 596
Db 1504 TEMEDLSSKDDVGKS---VHELEKSKRALEEQ--VEEMKTQLEEELEDELOATEDAKLRL 1558
Qy 597 CVPIQS 602
Db 1559 EVNLQA 1564

RESULT 10
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRN
; ORGANISM: Human
US-09-949-016-11433

Query Match 4.5%; Score 145; DB 4; Length 2733;
Best Local Similarity 19.4%; Pred. No. 0.0017;
Matches 134; Conservative 117; Mismatches 250; Indels 188; Gaps 33;

Qy 18 VKHVEKIKIE-----SCSERTKIFRLYSVCTVEKALILARONRQNKAVENLRSLT 70
Db 785 LKEIEAKVELELKVSGSTTSELTK-----KSBEVFQIQKQKQGLEIESL-KTVS 834
Qy 71 HALLMWGASYLFDKLDHPSSETPDGVSFQSGIMDGVHFEFSSILSSKGGEEVVKLCLE 130
Db 835 HEAEVHAES-LQKLE---SSQIQIAGLEHLRELQ-KLDELQKLSKK--BEDVSYLSG 887
Qy 131 LLBAKHAQGTYSDDSTLFGEDHTKL-----SDESPNIFWSKLLGGKQPMWKYPSDT 182
Db 888 QLSEKKAALTKIQTEIEQEDDLIKALHTQLEMQAKEHDERI---KQLQVELCEMKQKPEEI 945
Qy 183 PQNRKRQVY---FEGSEASPKTGDGNAKRRKASDDV-----TDPRTVTPPDV-DEER 233
Db 946 GEESRAKQIQKRLQAALTSRKEA---LKENKSLQEEELSARGTIERLTKSLADVESQV 1001
Qy 234 KASGKDQH---MGAL-----ESPKVIT-----LQSSCKS---SGTGTGLDGNDAFG 272
Db 1002 SAQNKEDTVGLRALLQOEERDKLITEMDRSLLENQSLSSCESLKLALSGLTEDKE--- 1058
Qy 273 LYSMGSHISGIPEDMLAS-----QDMGKIPDESQRRLHTV----- 307
Db 1059 --KLKVEIESLSKSIABESTEWQKHKELOKEYEILLQSYENVSNEAERIQHVVAVRQE 1116
Qy 308 -----LKPMAKLCQVLHLSDACTSMVGNFLEYV 336
Db 1117 KQELYGKLRSTANKKETEKQLOEABQEBEEMKMRKFAK-----SKQKQILELE 1167
```

QY 337 IENHRIYEE--PATTFOAFQIALSWIAALLVKQILSHKESLVRANSELAPKCSRVEDY- 393
Db 1168 EENDRLRAEVHPAGD-----TAKECH-----ETLSSNASMKBELRYKMEYE 1210
QY 394 -----IYSILSCMSKLFLEHTQGLQFCFGTNSKQSVVSTKLVNESLSGATVRDEKIN-- 446
Db 1211 TLSKKFQSLMSEKDSL-SEEVQDLKHQIEGNSVSKQA-----NLEATEKHNDQNTVT 1260
QY 447 ---TKSMRNSSEDEECM---TEKRCSHYSTATRDIEKTIISGKKYKK--QVQKLVQSH 498
Db 1261 BEGTQSIPEETEODLSMRPTCSSEVPSAKSANPAVS---KDFSSHDEINNYLQIID 1317
QY 499 EKKMELLNMYADKKQKLTSTSVSAVIRITCSRTSTQVGLKLLDHNHYERKFEIKSEK 558
Db 1318 QLKERIAGLEBEKQKNKFSOTLENEK-NTLLSQISTQGBELKMLQEEV-TMNLNLNOOI 1375
QY 559 NECLKSLRQMEHVAKKLAEDACWINRI 587
Db 1376 QEELSRVTKLKETAEEKDDLEERLMNQL 1404
RESULT 11
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: Curafatseqformatter Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154
Query Match 4.4%; Score 140.5; DB 4; Length 3210;
Best Local Similarity 21.0%; Pred. No. 0.0059;
Matches 149; Conservative 101; Mismatches 238; Indels 221; Gaps 37;
QY 7 LFGSSL-----NPSHDVKEKIKIESCBERTKIFRLYSVCTVEEKALILARQNM 57
Db 1622 VFCSLQENLUTRETPTAPAKGV--ELESCE---VYR-QSLEKEEK---MESQIM 1672
QY 58 QNKAVENLRLTHALLMWGASYLFDKLDHPHSSETPDGVSFQSGINDGVIHFSSILS 117
Db 1673 KNKEIQELEQ-----LLSSRQELDCRLKQVLSNEQ-----NQKLTSTVLEMSKLA 1721
QY 118 SKGSENEVKCLLLEAKHAGTYSSTLFG---EDHIKLSDESPNIFWSKLLGGKNP 174
Db 1722 AEKQTEQLSLEVARLQGLDLSRSLIGIDTDAIQ-GRNESCIDI--SK----- 1771
QY 175 MWKYPSPDPQNRXK-VQYFEGSGAS-----PKTGDGNNAKKRKKASDDVTPRVTD 226
Db 1772 --EHTSETETPHDVHQICDKARQQLNDIEKITGTAVKPTGCGSGEQSPDTNYPE 1829
QY 227 PVDDDERKASGDHMGALSPKVTITQSSCKSSGTDGLDGNDAFGLYMSGSHISGIPED 286
Db 1830 PGED---KTQG-----SSECIS---ELSFSGPNAL-----VPM 1857
QY 287 MLASQD-----WGKIPDESQRRLHTV-----LKPMAKLCQVHLHS----- 322

Db 1858 FLGNQEDIHNLQLRVKTSENENLRLHVIEDRRDRKVESLNLNEMKELDSKHLQEVQMTK 1917
QY 323 -DACTSM---YGNF-----LEVNIENHRIYEEPATTFQAFQIALSW----- 359
Db 1918 IEACIELEKIVGELKKNESDLSEKLEFSCDQHQELLQORVETSEGLNSDLEMHADKSSRED 1977
QY 360 IAAILLVQILSHKESLVRANSELA-FKCSRVEDYIYIISLSCMSKLFLE-----HTOGL 412
Db 1978 IGDNVAKVNDKSWKRFLDVENELRIIRSEKASIEH-----EALYLEADLEVVOQTEKL 2029
QY 413 QFDCFGTNSKQSVVSTKLVNESLSGATVR-----DEKINTKSMR 451
Db 2030 CLEKDNENKQVIV---CLBEEISVVTSERNQLRGELDTMSKTTALTDLQSEKMKETQE 2086
QY 452 NSSDEECM-----TE-----KRCSHYSTATRDIEK---TISGI 482
Db 2087 LESHQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQELQSLKSDQSALSIT 2146
QY 483 KKKYKKQVQKLVQSH--EKKMELLNMYADKK--OKLETSSKSVANVR-----ITCST 533
Db 2147 KCELENQIAQLNKEKELLVKESLQARLSSESDYEKLNVSKALEALVEKGEFALRLSST 2206
QY 534 STQVGDCLKLLDHNHYERKFDK--IKSEKNECLSKLEQMEHVAKKLAED 580
Db 2207 QEEV-----HOURRGIEKURVREADE-----KKQLHIAEKJERERE 2244
RESULT 12
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

```
Query Match      4.4%; Score 140.5; DB 1; Length 3248;
Best Local Similarity 21.0%; Pred. No. 0.006;
Matches 149; Conservative 101; Mismatches 238; Indels 221; Gaps 37;

QY 7 LFSSSL-----NPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNM 57
Db 1622 VFCSSLQENLTKRTPSAKGV--ELESCE--VYR-QSLEKLEEK--MESQIGIM 1672
QY 58 QNKAVENLRSLTHALLMAGASYLFDKLDHFSSETPDGSGVSEQSMGDIHVFHFSILS 117
Db 1673 KNKEIQELEG-----LLSSERQELCLRKQYLSENEQ-----WQKLTSTVLEMSKLA 1721
QY 118 SKGGEENEVKCLLLEAKHAQGYSSDSTLFG--EDHIKLSDEESPNIFMSKLLGKNP 174
Db 1722 AEKKQTEQLSELEVARLQGLDLSRLGIDTDAIQ-GRNESCDI--SK----- 1771
QY 175 MWKYPSTDPORNKR--VQYFEGSEAS-----PKTGDGNAKRRKASDDVDTPRVTDP 226
Db 1772 --EHTSETTTPKHVDHQICDKDAQDNLNLDIEKITETGAVKPTGECGSEQSPDTNYPE 1829
QY 227 PVDDERKAGKDHMGALSPKVTITLQSSKSGSTGDTLGDNDAFGLYSMGSHIGIPED 286
Db 1830 PGSD---KTQG-----SSSECIS--ELSFSGPNAL-----VPMD 1857
QY 287 MLASQD-----WGKIPDESORRLHTV-----LKPMAKLCQVLHLS----- 322
Db 1858 FLGNQEDIHNLQURVKETSNENLRLLHVIEDRRDKVESLNEKMLDLSKHLQEVQMTK 1917
QY 323 -DACTSM---VGNF-----LEYVIENHRIYEBPATTFOAFQIALSW----- 359
Db 1918 IEACIELEKIVGELKKNESDLSEKLYFSCDHQELLQRTVETSEGLNSDLEHAKSSRED 1977
QY 360 IAALLVKQILSHKESIVRANSELA--FKCSRVVDYIYISLCMKSFLFLE-----HTQGL 412
Db 1978 IGDVAKVNDWSKERFLDVENELSRIRSEKASIEH-----EALYLEADLEVVOQTEKL 2029
QY 413 QPDCFGTNSKQSVVSTKLVNESISGATVR-----DEKINTKSMR 451
Db 2030 CLSKDNENKQKIV---CLSEELSVVTSERNQURGELDTWMSKTTALDQJSEKMKETQE 2086
QY 452 NSSEDEECM-----TE-----KRCSHYSTATRDIEK---TISGI 482
Db 2087 LESHQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALSLT 2146
QY 483 KKKYKQVQKLVQEH--EKKMELLNMYADK--OKLETSKSVAAVIR-----ITCSRT 533
Db 2147 KCELENQIAQNKKEKELLVKESESLOARLSSESDYEKLVNKALEAALVEKGEFALRLSST 2206
QY 534 STQVGDCLKLDHNYVERKFE--IKSEKNECLKSLEOMHEVAKKLAED 580
Db 2207 QBEV-----HQRRGIEKLVRIBADE-----KKQLHIAEKLERERE 2244

RESULT 13
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street, Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match      4.4%; Score 140.5; DB 5; Length 3248;
Best Local Similarity 21.0%; Pred. No. 0.006;
Matches 149; Conservative 101; Mismatches 238; Indels 221; Gaps 37;

QY 7 LFSSSL-----NPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNM 57
Db 1622 VFCSSLQENLTKRTPSAKGV--ELESCE--VYR-QSLEKLEEK--MESQIGIM 1672
QY 58 QNKAVENLRSLTHALLMAGASYLFDKLDHFSSETPDGSGVSEQSMGDIHVFHFSILS 117
Db 1673 KNKEIQELEG-----LLSSERQELCLRKQYLSENEQ-----WQKLTSTVLEMSKLA 1721
QY 118 SKGGEENEVKCLLLEAKHAQGYSSDSTLFG--EDHIKLSDEESPNIFMSKLLGKNP 174
Db 1722 AEKKQTEQLSELEVARLQGLDLSRLGIDTDAIQ-GRNESCDI--SK----- 1771
QY 175 MWKYPSTDPORNKR--VQYFEGSEAS-----PKTGDGNAKRRKASDDVDTPRVTDP 226
Db 1772 --EHTSETTTPKHVDHQICDKDAQDNLNLDIEKITETGAVKPTGECGSEQSPDTNYPE 1829
QY 227 PVDDERKAGKDHMGALSPKVTITLQSSKSGSTGDTLGDNDAFGLYSMGSHIGIPED 286
Db 1830 PGSD---KTQG-----SSSECIS--ELSFSGPNAL-----VPMD 1857
QY 287 MLASQD-----WGKIPDESORRLHTV-----LKPMAKLCQVLHLS----- 322
Db 1858 FLGNQEDIHNLQURVKETSNENLRLLHVIEDRRDKVESLNEKMLDLSKHLQEVQMTK 1917
QY 323 -DACTSM---VGNF-----LEYVIENHRIYEBPATTFOAFQIALSW----- 359
Db 1918 IEACIELEKIVGELKKNESDLSEKLYFSCDHQELLQRTVETSEGLNSDLEHAKSSRED 1977
QY 360 IAALLVKQILSHKESIVRANSELA--FKCSRVVDYIYISLCMKSFLFLE-----HTQGL 412
Db 1978 IGDVAKVNDWSKERFLDVENELSRIRSEKASIEH-----EALYLEADLEVVOQTEKL 2029
QY 413 QPDCFGTNSKQSVVSTKLVNESISGATVR-----DEKINTKSMR 451
Db 2030 CLSKDNENKQKIV---CLSEELSVVTSERNQURGELDTWMSKTTALDQJSEKMKETQE 2086
QY 452 NSSEDEECM-----TE-----KRCSHYSTATRDIEK---TISGI 482
Db 2087 LESHQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALSLT 2146
QY 483 KKKYKQVQKLVQEH--EKKMELLNMYADK--OKLETSKSVAAVIR-----ITCSRT 533
Db 2147 KCELENQIAQNKKEKELLVKESESLOARLSSESDYEKLVNKALEAALVEKGEFALRLSST 2206
```

Qy 534 STQVGDLLKLDHNYERKDE--IKSEKNECLKSLEQMHVAVAKKLADE 580
Db 2207 QEEV-----HQLRRGIEKLVRLEADE-----KKQLHIAEKLKERERE 2244

RESULT 14

US-09-538-092-1044
; Sequence 1044, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 1044
; LENGTH: 1427
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)_(0)
; OTHER INFORMATION: Polypeptide Accession Number P30622
US-09-538-092-1044

Query Match 4.3%; Score 139.5; DB 4; Length 1427;
Best Local Similarity 20.7%; Pred. No. 0.0019;
Matches 124; Conservative 108; Mismatches 229; Indels 137; Gaps 28;
Qy 31 ERTKIFRLYSVCTVEEKALILARQNMQRQKAVENLRLTHALLMWGASYLFDKLDHPS 90
Db 570 EETHQKEIKALYATEK---LSKENESLKSLEHANKENSVDIALWKS-----KL----- 616
Qy 91 SETPDSGVSPQSIQMDGVHIFPSSILSKSGEENEVK-----ICLLLEAKHAQGYSS 143
Db 617 ---ETAISHQQAMEELKVSFGGLGTETAEFAELQTQIEKMLDYOHEIENLQNOQDS 672
Qy 144 DSTLFGED---HIKL-----SDESPNI FWSKLLGKNPMWKYPSTDPORNRK---RVQ 191
Db 673 ERAHAKEMEALRAKLMKVIKEKENSLEAIRSKDKAEDQHLVEMEDTLNKLQAEIKVK 732
Qy 192 YFEGSEASPKTGGGNNAKRRKASDDVTDPP-RVTDPPVDDDE--RKAS--GKDHMGAL 246
Db 733 ELEVLA-----KCNQETKVIDNFTSOLKATEEKLLDLALRKASSEKSEMKKLRLQ 784
Qy 247 P-----KVITLQ---SSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWGK 295
Db 785 QLEAAEQIKHLEIEKNAESSKASITRELQREL-----KLTLNQLNL-----SEVSQ 833
Qy 296 IPDESORLHTVLPKMAKLQVHLSDACTSMWGNPLEYVNIENHRIYEEPAITTFQAFQI 355
Db 834 VKETLEKELQ-ILKEKFAE-----ASEEAVSVORSQMOETWNKLHQKEEQ-----FNM 879
Qy 356 ALSWIAALLVQKILSHKESLVRANSELAFKCSRYEVDYIYISLCSMKSLFLEHTQGLQFD 415
Db 880 LSSDLEKL--RENLADEAKFREDEREQLIKAKEKLENDIAEIMK----- 924
Qy 416 CFGTNSQSVSTKLINP-SLSGATVRDEKIN-TKSMRNSSEDECEMTEKCSHYSTATR 473
Db 925 MSGDNSSQ---LTKMNDLRLKERDVEBELQKLTKANENAS----- 962
Qy 474 DIETISGIKKKYKQVKLQVHEEKKWELLNMYADKKOKLET5-----KSVEAAVIRIT 529
Db 963 FLQKSIDMTVKASQSQEAAKHEEKKLELRLKSLDEKKQWETSHNQCEKQELKARYERAT 1022

Qy 530 CSRTSTQVGDLL-----KLDHNYERKDEIKSEKNECLKSLEQMHVAVAKKLADEA 581
Db 1023 -SETKTRHEILQNLQKTLLD--TEDKLGARENSGLLQLEELRKOADKAKAQA 1077

RESULT 15

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 4.3%; Score 139.5; DB 1; Length 2482;
Best Local Similarity 21.0%; Pred. No. 0.0048;
Matches 149; Conservative 101; Mismatches 238; Indels 221; Gaps 37;
Qy 7 LFGSSL-----NPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNM 57
Db 894 VFCSSLOEENLTKRBTSPAKGVE--ELESLSCE---VYR-QSLEKLEEK---MESQGIM 944
Qy 58 QNKAVENLRLTHALLMWGASYLFDKLDHPSSETPDSGVSPQSIQMDGVHIFPSSILS 117
Db 945 KNKEIQLEQ-----LLSSERQELDCRLKQYLSENEQ-----WQQLTSTVLEMSKLA 993
Qy 118 SKGGEENEVKCLLLEAKHAQGYSSDSTLFG-----EDHKLSDSESPNIFWSKLLGKNP 174
Db 994 AEKQTEQLSLELEVARLQLQGLDLSRSLIGDITDAIQ--GRNESCDI--SK----- 1043
Qy 175 MWKYPSTDPORNRK-VQYFEGSEAS-----PKTGGGNNAKRRKASDDVTDPRVTD 226
Db 1044 --EHTSETTERTPKDGHQICDQAQDNLNDIEKITETGALKPTGSCGSEQSPDTNYPE 1101
Qy 227 PVDDDERKAGKDHMGALSPKVTITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPED 286
Db 1102 PGED---KTQG-----SSCLIS---ELSFSGPNAL-----VPM 1129
Qy 287 MLASQD-----WGKIPDESQRRLHTV-----LKPMAKLQVHL 322

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:58:28 ; Search time 57.6667 Seconds

(without alignments)
4160.324 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILGSLNPSHDVHKH.....SGNNKHFSGSNISQNPADV 621

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1726218 seqs, 386331768 residues

Total number of hits satisfying chosen parameters: 1726218

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3207	99.7	2254	17 US-10-732-923-18943	Sequence 18943, A
2	590.5	18.4	1990	16 US-10-437-963-178706	Sequence 178706, A
3	495.5	15.4	2630	16 US-10-437-963-172374	Sequence 172374, A
4	456.5	14.2	1954	16 US-10-437-963-178698	Sequence 178698, A
5	335.5	10.4	358	16 US-10-425-115-343695	Sequence 343695, A
6	192.5	6.0	162	15 US-10-424-599-155808	Sequence 155808, A
7	163.5	5.1	1461	17 US-10-732-923-8689	Sequence 8689, A
8	163.5	5.1	1465	15 US-10-369-493-5100	Sequence 5100, A
9	163.5	5.1	1465	17 US-10-732-923-8690	Sequence 8690, A
10	162.5	5.1	1739	15 US-10-341-434-61	Sequence 61, Appl
11	162.5	5.1	1739	17 US-10-732-923-8832	Sequence 8832, A

12	161	5.0	1445	17 US-10-732-923-8150	Sequence 8150, A
13	160	5.0	1738	17 US-10-732-923-8335	Sequence 8335, A
14	158	4.9	1015	15 US-10-369-493-3210	Sequence 3210, A
15	157	4.9	126	16 US-10-425-115-329771	Sequence 329771, A
16	152	4.7	2552	16 US-10-437-963-129734	Sequence 129734, A
17	150.5	4.7	1358	15 US-10-675-072A-43	Sequence 43, Appl
18	150	4.7	1785	17 US-10-732-923-8672	Sequence 8672, A
19	148	4.6	699	16 US-10-425-115-293739	Sequence 293739, A
20	147.5	4.6	728	17 US-10-732-923-8570	Sequence 8570, A
21	147.5	4.6	1813	17 US-10-732-923-8155	Sequence 8155, A
22	147.5	4.6	1959	15 US-10-028-248A-36	Sequence 36, Appl
23	147.5	4.6	1959	15 US-10-107-782-36	Sequence 36, Appl
24	146	4.5	1881	14 US-10-032-585-7646	Sequence 7646, A
25	145.5	4.5	1647	14 US-10-032-585-7572	Sequence 7572, A
26	145.5	4.5	6761	17 US-10-732-923-15035	Sequence 15035, A
27	145	4.5	1960	15 US-10-236-031B-62	Sequence 62, Appl
28	145	4.5	1960	15 US-10-028-248A-104	Sequence 104, A
29	145	4.5	1960	15 US-10-107-782-104	Sequence 104, A
30	145	4.5	1960	17 US-10-696-909A-49	Sequence 49, Appl
31	145	4.5	3225	16 US-10-408-765A-254	Sequence 254, A
32	143	4.4	1130	15 US-10-369-493-6751	Sequence 6751, A
33	143	4.4	1379	14 US-10-205-219-5	Sequence 5, Appl
34	142.5	4.4	1422	17 US-10-732-923-8665	Sequence 8665, A
35	142.5	4.4	1961	15 US-10-028-248A-103	Sequence 103, A
36	142.5	4.4	1961	15 US-10-107-782-103	Sequence 103, A
37	142	4.4	1426	17 US-10-732-923-8310	Sequence 8310, A
38	141	4.4	1074	16 US-10-437-963-102819	Sequence 102819, A
39	141	4.4	1806	17 US-10-732-923-8671	Sequence 8671, A
40	140.5	4.4	1643	17 US-10-732-923-8217	Sequence 8217, A
41	140.5	4.4	3114	16 US-10-751-736-99	Sequence 99, Appl
42	140	4.4	865	17 US-10-732-923-7041	Sequence 7041, A
43	140	4.4	906	17 US-10-732-923-6837	Sequence 6837, A
44	139.5	4.3	1829	17 US-10-732-923-8687	Sequence 8687, A
45	139.5	4.3	3113	16 US-10-723-860-862	Sequence 862, A

ALIGNMENTS

RESULT 1

US-10-732-923-18943
; Sequence 18943, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18943
; LENGTH: 2254
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-18943

Query Match	99.7%	Score 3207;	DB 17;	Length 2254;
Best Local Similarity	99.7%	Pred. No. 3.3e-237;		
Matches 619;	Conservative	1;	Mismatches	1;
Indels	0;	Gaps	0;	
Qy	1	RADAFILGSLNPSHDVHKHVKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQRNK	60	
Db	883	RADAFILGSLNPSHDVHKHVKIKIESCSERTKIFRLYSVCTVEEKALILARQNRQNK	942	
Qy	61	AVENLNRSALTHALMWCASYLFDKLDHFHSSETPDGSVFQSIQMDGVIHFSSILSSKG	120	
Db	943	AVENLNRSALTHALMWCASYLFDKLDHFHSSETPDGSVFQSIQMDGVIHFSSILSSKG	1002	
Qy	121	GEENEVKKLLLEAKHAQGYSSDSTLFGEDHIKLSDEESPNIFWSKLLGKGNPMWKYPS	180	
Db	1003	GEENEVKKLLLEAKHAQGYSSDSTLFGEDHIKLSDEESPNIFWSKLLGKGNPMWKYPS	1062	

```
QY 181 DTQRNRKRVQYFEGSEASPKTGDGNAKKRKKASDDVTPRVTDPPVDDDERKASKOH 240
Db 1063 DTQRNRKRVQYFEGSEASPKTGDGNAKKRKKASDDVTPRVTDPPVDDDERKASKOH 1122
QY 241 MGALESFKVITLQSSCKSSGTDGLDGNDAFGLYSMSGSHISGIPEDMLASQDWGKIPDES 300
Db 1123 MGALESFKVITLQSSCKSSGTDGLDGNDAFGLYSMSGSHISGIPEDMLASQDWGKIPDES 1182
QY 301 QRLHVTLPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRIYEEPPATTFQAFQIALSWI 360
Db 1183 QRLHVTLPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRIYEEPPATTFQAFQIALSWI 1242
QY 361 AALLVKQILSHKESLVRANSELAFKCSRVEDVITYLSLSCWKSILFLEHTQGLQDFDCGTN 420
Db 1243 AALLVKQILSHKESLVRANSELAFKCSRVEDVITYLSLSCWKSILFLEHTQGLQDFDCGTN 1302
QY 421 SKQSVVSTKLVRNLSGATVRDEKINTKSMRNSSEDECKTEKCSHYSTATRDIEKTIIS 480
Db 1303 SKQSVVSTKLVRNLSGATVRDEKINTKSMRNSSEDECKTEKCSHYSTATRDIEKTIIS 1362
QY 481 GIKKKYKQVQKLVQEHHEKKMELNNYADKKQKLETSKSVAAVIRITCSRTSTQVGD 540
Db 1363 GIKKKYKQVQKLVQEHHEKKMELNNYADKKQKLETSKSVAAVIRITCSRTSTQVGD 1422
QY 541 KLJDHNYERKFDTEKSEKCLSLQEHHEVAKKLADEACWNRKSKAAKLKVCVPI 600
Db 1423 KLJDHNYERKFDTEKSEKCLSLQEHHEVAKKLADEACWNRKSKAAKLKVCVPI 1482
QY 601 QSGNNKHFSGSSNISQNPADV 621
Db 1483 QSGNNKHFSGSSNISQNPADV 1503
```

RESULT 2

```
US-10-437-963-178706
; Sequence 178706, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 178706
; LENGTH: 1990
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76236C.1.pep
US-10-437-963-178706
```

```
Query Match 18.4%; Score 590.5; DB 16; Length 1990;
Best Local Similarity 26.2%; Pred. No. 8.5e-36;
Matches 175; Conservative 127; Mismatches 239; Indels 127; Gaps 20;

QY 3 DAFILFGSSLNPSHDVKKHVEKIKIESCSERTKIFRLYSVCTVEBEKALILARQNMQRNKA 62
Db 755 DALLIYCSDNWPTNDLVLQRISESQSECVPIFRLLYSCTVEBEKILAKHDHILDSNV 814
QY 63 ENLNRSLTHALLMGASYLFDKLDPHF-----SSETPDGVSVPFQSIWDGVIHFPSILS 117
Db 815 QNVMPVISHSILSNGASFLNRLBEFQKHQYSDKSDSDGL-----FMNNVFLFAAKLS 869
```

```
QY 118 SKGGEENEVKLCILLLEAKHAQTYSSDSTLFG- DHIKLSDBESPNI--FWSKLLGGKPN 174
Db 870 TNVEASTKMNVAISRAQSGSFYRDIAVISEREGISAVDGLPKFWFWFWSNLGGRSP 929
QY 175 MWKYPSTQPNRNRVQYFEGSEASPKTGDGNAKKRKKASDDVTPRVTDPPVDDDERK 234
Db 930 HMQYISEPVQRNRKIQNMEDQMRIPAEETDEAIMKRRKIGE-----IMD----- 974
QY 235 ASKGDHMGALSPKVVITLQSSCKSSGTDGLDGNDAFGLYSMSGSHISGIPEDMLASQD-- 292
Db 975 -----SSPKILPVKD-----NDAV-----LPENSTASSHE 1000
QY 293 -----WGKIPDES-----QRLHVTLPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRI 342
Db 1001 TSVDVDTWQELGAESLQGTQGLHTQLKPELSKLYELLELPETVKYCLCEBELLDVILKNHQV 1060
QY 343 YEPATTFQAFQIALSWIAALLVKQILSHKESLVRANSELAFKCSRVEDVITYLSLSCWK 402
Db 1061 SQEPKGIHAFNIALCWRAASLLKHKNRRESLALSVRLNLYECDEVLAELYVVEKLRILK 1120
QY 403 SLFLEHT-----QGLQDFCGTNSKQSVVSTKLVRNLS-LSGATVRDEKINTKSMRNSSE 455
Db 1121 KKFSRRASETSKOSQSTPVNNTSSYKQQTSPKLRSDGSIQCHQVTTTIDGLENVSHBEAPH 1180
QY 456 D-----ECCMTEKRCSHYSTATRDIEKTIISGIIKKYKQV--QKLVQEHHEKKMELNNYAD 510
Db 1181 DIUTEEMILEQK-----ELISVLETHREHVLRLDELLERITEKRINLINMVFS 1228
QY 511 KKQK-----LETSKSVAAVIRITCSRTSTQVGDLLKLJDHNYERKFDTEKSEK 558
Db 1229 LREKNTQDKQNETTLLDMHKQKQVAKLQBTCTN-----LVVEHLRKGHID--SEDR 1277
QY 559 NECLKSLQEHHEVAKKLADEACWNRK-----SMAAKLKVC-VPIQSGNNKHFSGSS 612
Db 1278 DATVKLIIEFWITLLYAFNLNMRCHQNKLMQOSTSWNKELQKLEIFLQQAESGHLDRSF 1337
QY 613 NISQNPAD 620
Db 1338 DQOIPLPD 1345
```

RESULT 3

```
US-10-437-963-172374
; Sequence 172374, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172374
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70516C.1.pep
US-10-437-963-172374
```

```
Query Match 15.4%; Score 495.5; DB 16; Length 2630;
Best Local Similarity 25.2%; Pred. No. 2.6e-28;
Matches 176; Conservative 108; Mismatches 255; Indels 159; Gaps 20;

QY 3 DAFILFGSSLNPSHDVKKHVEKIKIESCSERTKIFRLYSVCTVEBEKALILARQNMQRNKA 62
```

Db 881 DAIILYSDNNPLNDLAKALQKIKIESOFERVISIFRLYTPFTVEKSLVLAQGVINDNI 940
Qy 63 ENLNRSLTHALLMWGASYLFDKLDHFASSTPDGVSFQESIMDGVIEHFSILLSSKGE 122
Db 941 QDLTSLKXSLRWGAFLSRLDEVOQDDHASKSEMERHFIDEVIEFLTKSTVED 1000
Qy 123 ENVKLLCLLEAKHAQGYSDSTLFGEDHKLSDRESNFIWKLGLGKNPMWKPST 182
Db 1001 STEVHRKSIKANSNGELYSRNIITLMGEKISVLEDPNPAEFWLNLDGRSPHVCISEP 1060
Qy 183 PORNKRKVOYFEGSEAPKTDGNGNAKRRKASDDV---TDPRTDPPVD---DERKAS 236
Db 1061 LQSRVTSQTMW---EVNAPAEINEARKKRRKVGIMGSSKVSVDKSNDDALPDICTTS 1118
Qy 237 GKDHMGALSPKVTITLQSSCKSGTDLGNDAGFLYSMGSHISGIPEDMLASQDWGKI 296
Db 1119 GP---ALQ-PVDVTQKQSGESGSLM-----1142
Qy 297 PDSQRRLHVLKPMKAKLCQVLHLS-----DACTSMVGNFLEXVIE 338
Db 1143 --STPKLHAQMKQELSKLIKVLQLPNTIQFGAKIAGQVLI EVDNVTLLVQEFYLLN 1200
Qy 339 NHRIVESPATTFQAFQALSIAALLVKQILSHKESLVANSSELAFKCSRVEDVYIYSL 398
Db 1201 NHVVQEPKYIFHALNIALCWRVASIHFVVDHKSALAEKRLUKYECNEBELARLVYDSL 1260
Qy 399 SCMSLFLHTQGLQDFCGFNSKQSVVSTKLVNESLSGATVRDEKINTK-----448
Db 1261 ---KRPKPKAGATGSCQST---SVEKTKPSQOETSN-ILRNDHIFPKQMDLHDFM 1312
Qy 449 -----SMRNSDEBECM---TEKRCSHYSTATRD---IBKTSIGIKKKYKQVK 492
Db 1313 NGALQESFVAQMVSEQELIAPVGTMEC-HFSTDELPIVEKRLINDVNFVSREYR 1371
Qy 493 LVQHEEKWELLNMYADKKQKLETSKSVRAAVRITCSRTSTQVGLKLDHNYERKFD 552
Db 1372 IFDQSQSISELEKYQNTKRLAT-----VCN-----LVLEHTCRSHAD 1411
Qy 553 EIKSEKNECKLSLQMEHVAKKLAEDACWINRIKS-----WA-----AKLKVCVPIQSG 603
Db 1412 --VETRNDTIQTQVWFTMLMYAFLEHMRLOHSLQSLQNTWAEERQLKLEAKSG 1469
Qy 604 -----NNKHFSGSS 612
Db 1470 QLDHTFDQIAPDSNFMVQBFHLEKQSSNSHVSGSA 1507

RESULT 4

US-10-437-963-178698
; Sequence 178698, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 178698
; LENGTH: 1954
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1.pap

US-10-437-963-178698

Query Match 14.2%; Score 456.5; DB 16; Length 1954;
Best Local Similarity 24.0%; Pred. No. 1.7e-25;
Matches 160; Conservative 118; Mismatches 229; Indels 159; Gaps 24;
Qy 3 DAFILFSSSLNPSHDVKKHVEKIKIESCSERTKIFRLYISVCTVEKAILARQNRQKAV 62
Db 755 DAIILYSDNNPTNDRVLQRIESQSECVPIFRLYSSCTVEBKTLIL-----803
Qy 63 ENLNRSLTHALLMWGASYL---FDKLDHFASSTPDGVSFQESIMDGVIEHFSILLSSK 119
Db 804 ----RSM--IIFLTATSRILEEFQKHD-YSKSDESDGL-----FMNVFLEFAKLSTN 851
Qy 120 GGEENEYKLLCLLEAKHAQGYSDSTLFGEDHKLSDRESNFI--FWSKLLGKNPMW 176
Db 852 VEASTKMNVAISRAQSGSFYSRDIAVISEREGISAVDGLPKFWTFWSNLLGGRSPHW 911
Qy 177 KYPSDTPQNRKRVOYFEGSEAPKTDGNGNAKRRKASDDVTDPRVTDPPVDDDERKAS 236
Db 912 QYISEPVQRRRKIQNMEDQMRIPABETDEAIMKRRKIGE-----IMD-----954
Qy 237 GKDHMGALSPKVTITLQSSCKSGTDLGNDAGFLYSMGSHISGIPEDMLASQD---292
Db 955 -----SFPKILPVKD-----NDAV-----LPERSTASSSHETS 982
Qy 293 ---WGKIPDES---QRRRLHVLKPMKAKLCQVLHLSDACTSMVGNFLEXVIEHRIYE 344
Db 983 VDDTWQELGAESLQGTQKGLHTQLKPELSKLYELLELPETVKCLCEBLLDYILKNHQ---1039
Qy 345 EPATTFQAFQALSIAALLVKQILSHKESLVANSSELAFKCSRVEDVYIYSLCKMSL 404
Db 1040 -----CWRAASLLKHKINRRSLSALSVRLNLTANECDEVLAEYVVEKRLILKK 1086
Qy 405 FLEHT-----QGLQDFCGFNSKQSVVSTKLVNES---LSGATVRDEKINTKSMNSSED-456
Db 1087 PSRRASSTKQSGSTPVNNTSSYKQSTSPKRSRSGSICHQVTTIDGLENVSHAEAPHDI 1146
Qy 457 --BECMTKRCSHYSTATRDIEKTSIGIKKKYKQV---QKLQVHEEKWELLNMYADKK 512
Db 1147 LTEEMILEQK-----ELISVLETHREHVLRLDELLERITEKRLINLNMVFSLR 1194
Qy 513 QK-----LETSKSVRAAVRITCSRTSTQVGLKLDHNYERKFDKSEKNE 560
Db 1195 EKNIQDKQGNETTLDMHKKQEVAKLQETCN-----LVVEHLRKHGID--SEDRDA 1243
Qy 561 CLKSLEQMEHVAKKLAEDACWINRIK-----SWAAKLKVC-VPIQSGNNKHFGSSNI 614
Db 1244 TVKLITIEWFTLLLYAFLNHMRQHNKLMQOSTSWNKELQLKEIFLQQAASGHLDKRSFDQ 1303
Qy 615 SQNAPD 620
Db 1304 QIPLPD 1309

RESULT 5

US-10-425-115-343695
; Sequence 343695, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 343695
; LENGTH: 358
; TYPE: PRT

```
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)...(358)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_76611C.1.pep
US-10-425-115-343695

Query Match      10.4%; Score 335.5; DB 16; Length 358;
Best Local Similarity 31.0%; Pred. No. 3.2e-17;
Matches 102; Conservative 58; Mismatches 120; Indels 49; Gaps 12;

Qy 3 DAFILFSSNPSHDVKRVKIKIESCSERTKIFRLYSVCTVEEKALILARONRNKAV 62
Db 47 DALLIYCSNDPNWDLRALQRVSMESQSEVPFRLYSSTVEEKALILAKHDLSDNI 106
Qy 63 ENLNRSLTHALLMWGASYLFDKLDHF--HSETPDGVSFEQSIMDGVIIHFSSILSKG 120
Db 107 VNITPSLSHLLSWGASFLFNRLLEELQOHSY---SNVSGDELFMNDVLEFLTKLLSKV 162
Qy 121 GENEVUKLLEAKHAQGYSSDSTLFGF-DHKLSDSESPNI--FWSKLGGKPNMWK 177
Db 163 ELRTESGNTAISQALYCGSPFYRAIVVAGERGIPSYDGLPKFWAYVWSLLNGRSPQW 222
Qy 178 YPSDTPORNKRKVQYFEGSEASPKTGGGNAKRKKASDDVTDPRVTDPPVDDDERKASG 237
Db 223 YISEPVQSRRKI---NNMQQLKNTD----KQLKITETDEARV-----KRRIG 267
Qy 238 K-DHMGALSPKVITLQSSCKSGTDTGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKI 296
Db 268 EIMDSSAIDSP-----GKNDITILPGNNT---PPSSH-----QISVEDTWQEL 307
Qy 297 P-----DESQRRLHTVLKPKMAKLCQVHL 321
Db 308 ERSNLHATQKGLHVQLKPEISKLYKLLQL 336

RESULT 6
US-10-424-599-155808
; Sequence 155808, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 155808
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1.pep
US-10-424-599-155808

Query Match      6.0%; Score 192.5; DB 15; Length 162;
Best Local Similarity 29.2%; Pred. No. 1.1e-06;
Matches 57; Conservative 32; Mismatches 71; Indels 35; Gaps 4;

Qy 337 IENHRIYEEPAATFQAFQIALSWTAALLVKQILSHKESLVANSSELPKCSRVEVDYIYS 396
Db 1 MNNLYVNRPEPESIQAFQSLCWTAAASLLKHKLDPPIASLIQ---DLNFECKKEVDYICS 57
Qy 397 ILSCKMSLFLEHTQGLQFDCFGTNSQSVSTKLVNSLSGATVRDEKINTKSNRNSSED 456
Db 58 MLCKLKIFLYRTGNHYDTGF-----TKASGSPSRA 88

/ ORGANISM: Caenorhabditis elegans
/ TYPE: PRT
/ ORGANISM: Caenorhabditis elegans
US-10-732-923-8689

Query Match      5.1%; Score 163.5; DB 17; Length 1461;
Best Local Similarity 17.6%; Pred. No. 0.0042;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

Qy 2 ADAPILFGLSPLNSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARONRNQKA 61
Db 799 ADTVIIFDSNDPNQDLQMSRAHRIGQTKTVNIYRLVTGKSVEE-----EI 845
Qy 62 VENLNRSLTHALLMWGASYLFDKLDH-----FHSSETPDGVSFEQSIMDGV 108
Db 846 VERAKRKL-----LDHLVIQRMDDTGTVLTKNATASGVPFDKQBSLAI 891
Qy 109 IHEFSSI--LSSKGGENEV-----KCLLLEAKHAQGYSSDSTLFGF-DHKLSDSES 160
Db 892 L-KFGAVELFKKEGEEQPEVDIDILMGAETREAEVEYKENVKELLNLSFKYANFAIDEE 950
Qy 161 PNIF-----WSKLLGGKPNMWKYPSPDTPQNRKRV-----Q 191
Db 951 KDTAAATDEWAAI-----PEEDNRILLEERMKELAEWNLAPROKQPIQ 997
Qy 192 YFEGSEA-SPKTGDGNGNAKKRKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRFKFSKFMPLNRLLEEIAQDAELE 1057
Qy 235 ASGKHMGALSPKVITLQSSCKSGTDTGLDGNDAFLGYSMGSHISGIPEDMLASQDWG 294
Db 1058 EHSTDEMCKL-----VBSLSEACKKADE--FDSNEKNG-----DAG 1092
Qy 295 KIPDESQ-----RRLHTVLKPKMAKLCQVHLSDACTSWGVNFLEYV 336
Db 1093 AAESEKKDTERKPFHTCDVNLKQIERSHAELK-----LHEILKSETKTS----- 1139
Qy 337 IENHRIYEEPAAT-----FQAFQIAL-----SWIAALLVKQI-LSHKESL 375
Db 1140 -----FKPPANAKLQGWDDVMSRPDDSAALLGVWKYGYGSWEAIKMDPTGLADKIFI 1193
Qy 376 VRANSELAFKCSRVVDYIYSILSCMSLFLFHTQGLQFDCFGTNSQSVSTKLVNSL 435
Db 1194 KDKTKPQGNLQVRVDYLLKLM-----SKDKVKTTE----- 1225
Qy 436 SGATVRDEKINTKSNRNSSEDECEMTEKCRSHYSTATRDIEKTIISGKKYKKQVQLVQ 495
Db 1226 -----KKERKKADDVPVGEKKKGRHNNVPOGEK-----KKKEK----- 1262
Qy 496 EHEKXKMLNMYADKKQLKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKPFDEIK 555
```

Db 1263 --EKNSSSLK---DQLALLSIDKSLYGGALEDSAKPFLBCVKLCMPVHKMKLKEAQ 1317
Qy 556 SEKNECLKS-----LEQMHEVAKKLADEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKQADEAKYLTRLGDSFLENLETLIKPKTKIRKWNLYLWIFLCKFTLREP 1371
RESULT 8
US-10-369-493-5100
; Sequence 5100, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5100
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5100

Query Match 5.1%; Score 163.5; DB 15; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0042;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;
Qy 2 ADAFILGSSINPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEKALILARQNRQKA 61
Db 799 ADTVIIFDSWNPNQDLQAMRAHRIGQTKVTNIYRLVTKGSVEE-----EI 845
Qy 62 VENLNRSILTHALLMWGASYLFDKLDH-----FHSSETPDSGVSFQESIMDGV 108
Db 846 VERAKRKL-----LDHLVIQRMDDTGTGLVSKNATAGSVFDFQELSAI 891
Qy 109 IHEFSSI--LSSKGEENEV-----KCLLLEAKHAQGYSSDSTLFG-DEHIKLSDEES 160
Db 892 L-KFGAVELFKKEGEQEPEVDIDRLMGAETREABEEVMEKENVKELLSSFKYANFAIDEE 950
Qy 161 PNIF-----WSKLGKKNPMKYPSDTPQRNRKV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDNRILBEERMKELAEMLAPRQRKQPIQ 997
Qy 192 YFEGSEA-SPTGCGNAKKRKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDGDDEEDDTGKKKKKAVGNFTIPEIKRFIKSPKFSPLNRLBEIAQDALE 1057
Qy 235 ASGKHDMGALSPKVITLQSSCKSSGTDGLDNDAGFLYSGMHSIGIPEDMLASQDWG 294
Db 1058 EHSTDMEKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
Qy 295 KIPESQ-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYV 336
Db 1093 AAESEKKDIERKFKFHTCDVNLKQIERSHAELKP-----LHEILKSEETKTS-----1139
Qy 337 IENHRIYEPAATT-----FQAFQIAL-----SWIAALLVKQI-LSHKESL 375
Db 1140 -----FKPPANAKLQKGDVDMRSPDSSALLGVWKYGYGSWEAIKMDPTGLADKIFI 1193
Qy 376 VRANSELAFKSRVEVDYIYSILSCMSLFLFHTQGLQFDCFGTNSKOSVSTKLVBESL 435
Db 1194 KDKTKKPGKNLQVRVDYLLKLM-----SKDKVKTTE-----1225
Qy 436 SGATVRDEKINTKSMRNSSEDEECWTEKCRSHYSTATRDIEKTIIGIKKKYKQVQKLQV 495

Db 1226 -----KKRKRKADDDVPVGPPEKKKXKXHTNNVQEGEK-----KKKXKK-----1262
Qy 496 EHEBKRMELLMYADKKQKLETSKSVBAAVIRITCSRTSTQVGDKLKLDHNYERKFDIBK 555
Db 1263 --EKNSSSLK---DQLALLSIDKSLYGGALEDSAKPFLBCVKLCMPVHKMKLKEAQ 1317
Qy 556 SEKNECLKS-----LEQMHEVAKKLADEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKQADEAKYLTRLGDSFLENLETLIKPKTKIRKWNLYLWIFLCKFTLREP 1371
RESULT 9
US-10-732-923-8690
; Sequence 8690, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8690
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-732-923-8690

Query Match 5.1%; Score 163.5; DB 17; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0042;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;
Qy 2 ADAFILGSSINPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEKALILARQNRQKA 61
Db 799 ADTVIIFDSWNPNQDLQAMRAHRIGQTKVTNIYRLVTKGSVEE-----EI 845
Qy 62 VENLNRSILTHALLMWGASYLFDKLDH-----FHSSETPDSGVSFQESIMDGV 108
Db 846 VERAKRKL-----LDHLVIQRMDDTGTGLVSKNATAGSVFDFQELSAI 891
Qy 109 IHEFSSI--LSSKGEENEV-----KCLLLEAKHAQGYSSDSTLFG-DEHIKLSDEES 160
Db 892 L-KFGAVELFKKEGEQEPEVDIDRLMGAETREABEEVMEKENVKELLSSFKYANFAIDEE 950
Qy 161 PNIF-----WSKLGKKNPMKYPSDTPQRNRKV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDNRILBEERMKELAEMLAPRQRKQPIQ 997
Qy 192 YFEGSEA-SPTGCGNAKKRKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDGDDEEDDTGKKKKKAVGNFTIPEIKRFIKSPKFSPLNRLBEIAQDALE 1057
Qy 235 ASGKHDMGALSPKVITLQSSCKSSGTDGLDNDAGFLYSGMHSIGIPEDMLASQDWG 294
Db 1058 EHSTDMEKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
Qy 295 KIPESQ-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYV 336
Db 1093 AAESEKKDIERKFKFHTCDVNLKQIERSHAELKP-----LHEILKSEETKTS-----1139
Qy 337 IENHRIYEPAATT-----FQAFQIAL-----SWIAALLVKQI-LSHKESL 375
Db 1140 -----FKPPANAKLQKGDVDMRSPDSSALLGVWKYGYGSWEAIKMDPTGLADKIFI 1193
Qy 376 VRANSELAFKSRVEVDYIYSILSCMSLFLFHTQGLQFDCFGTNSKOSVSTKLVBESL 435
Db 1194 KDKTKKPGKNLQVRVDYLLKLM-----SKDKVKTTE-----1225
Qy 436 SGATVRDEKINTKSMRNSSEDEECWTEKCRSHYSTATRDIEKTIIGIKKKYKQVQKLQV 495

Db 1226 -----KKERKKAADVPVPGPEKKKRRHTNNVPQEGEK-----KKKEKK-----1262

Qy 496 EHEKQKELLNMYADKKQKLETSKSVENAVRITCSRTSTQVGLKLLDHNHYKPFDEIK 555

Db 1263 --BEKNSSSLK---DQALLSIDKSLYGGALDSSAKPFLCVCVLCMPVHKYMKKLEAQ 1317

Qy 556 SEKNECLKS-----LEQMEHVAKKLADEACWINRIKSWAAKLKVCVP 599

Db 1318 EAKNQADEAKYLRGLDSFLENLETLIKKPKTNIRKWNLYWLPFLCKFTLREP 1371

RESULT 10

US-10-341-434-61

Sequence 61, Application US/10341434

Publication No. US20030215835A1

GENERAL INFORMATION:

APPLICANT: Origene Technologies

TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes

FILE REFERENCE: 9U 204 205 R1

CURRENT APPLICATION NUMBER: US/10/341,434

CURRENT FILING DATE: 2003-07-18

PRIOR APPLICATION NUMBER: US 60/348,164

PRIOR FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: US 60/348,119

PRIOR FILING DATE: 2002-01-15

NUMBER OF SEQ ID NOS: 238

SOFTWARE: PatentIn version 3.1

SEQ ID NO 61

LENGTH: 1739

TYPE: PRT

ORGANISM: Homo sapiens

US-10-341-434-61

Query Match 5.1%; Score 162.5; DB 15; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0064;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

Qy 2 ADAPILGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEKA-----L 49

Db 879 ADTVVIFDSWNPQNDLQQAARHIGQKKQVNIYRLVGTGTVEEIIERAKKKQVLDHL 938

Qy 50 ILARQNQRONKAVEN-----LNRSLLTHALLMWSYLFDKLDHFSSETPDSGVSF 100

Db 939 VIQMDTTGRTILENNSGRSNPNFNKEELTAILKFGAEDLFKELE-----GEES 988

Qy 101 EQSIMDGVIIHFSSILSSKGGEENEVKLLLEAKHAQGTYSSTLFGEDHKLSDSES 160

Db 989 EQQEMD-----IDELRLAETRENEVSTSATDELL--SQFKVANPATMEDEEEL-----EER 1038

Qy 161 PNIFWSKLLGKNPMKYPSTPQRNKRQVYFE-----GSEASPTG 203

Db 1039 PHKMDWELI-----PEQRKKVEEERQKELEIYMLPRIRSTKKAQNTDS 1085

Qy 204 DGNNAKRRKASDDVTPDPRVTPDDDERKASGDHMGALSPKVITLQSSCKSSGTDG 263

Db 1086 DSDTESKQQRSSASESETED---SDDDKK-----PKRRGRPRSRVKDLVEG 1130

Qy 264 TLQGN-----DAFLYSGMHSIGIPEDML--ASQDWGKIPDESQRRLHTVLPKMAKLC 316

Db 1131 FTDAEIRRFIKAYKFF-----GLPLERLECLARD--AELVDKSVADL-----KRLG 1174

Qy 317 QVHLSDACTSMVGNFLEVYENHRIYEEBATTFOAFQIALSWIAALLVKQILSHKESLV 376

Db 1175 ELTH--NSCVSAMQYEEQKENASEGKGPQKR--RGPTIKISGV--QNVKSIHQHEEFE 1230

Qy 377 RANSEL-----AFKCSRVEVDYIYILSCMKSLFLEHTQGLQFCFGTNSK 422

Db 1231 MLHKSI PVDPEEKKKYCLTCRVKAAHFDVWGVGDDSLLLGIYEHGVG--NWELIKTDPE 1289

Qy 423 QSVVSTKLNESLSGATVRDEKINTKSMRNSSE-----DEECWTEKR 464

Db 1290 -----LXLTDKILPVET--DKKPOGQQLQTRADYLLKLLRKGLEKKGAVTGGEAKLKR 1342

Qy 465 -----CSHYSTATRDIEKTISGIIKKYKQVQKLVQEHBEKK 501

Db 1343 KPRVKKENKVPRLKEEHGIELSSPRSDNPSESGEVKDDGLEKSPMKKKQK--KKENKENK 1401

Qy 502 MELLNMYADK-----KQKLETSKSVENAVRITCSRTSTQVGLD---539

Db 1402 EKQMSRKQKQEGDKERKKSKDKKEKPKSGDAKSSKRSQGPVHITAGSEPPVIGDEDED 1461

Qy 540 -----LKLLOH-----NYERKFD-----EIKSEKNECLKS--564

Db 1462 DDLQDQTFSTICKERMPVKALKQLOKDPDKGLNVQSOLEHTRNCLLIKIGDRIAECLKAYS 1521

Qy 565 -----LEQMEHVAKKLADEACWINRIKSWAAKLKVCV 598

Db 1522 DQEHIKLWRRNLWIFVSKTFEPDARKLHLKYKWAHKRSQEEF-----SQKKKDDV 1572

Qy 599 PIQSGNNKHP-----SGSSN---ISQN 617

Db 1573 ---TGGKKPFRPEASGSSRDSLISQS 1595

RESULT 11

US-10-732-923-8832

Sequence 8832, Application US/10732923

Publication No. US20050108791A1

GENERAL INFORMATION:

APPLICANT: Edgerton, Michael D

TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

FILE REFERENCE: 38-15(52796)C

CURRENT APPLICATION NUMBER: US/10/732,923

CURRENT FILING DATE: 2003-12-10

PRIOR APPLICATION NUMBER: 10/310,154

PRIOR FILING DATE: 2002-12-04

NUMBER OF SEQ ID NOS: 24149

SEQ ID NO 8832

LENGTH: 1739

TYPE: PRT

ORGANISM: Homo sapiens

US-10-732-923-8832

Query Match 5.1%; Score 162.5; DB 17; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0064;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

Qy 2 ADAPILGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEKA-----L 49

Db 879 ADTVVIFDSWNPQNDLQQAARHIGQKKQVNIYRLVGTGTVEEIIERAKKKQVLDHL 938

Qy 50 ILARQNQRONKAVEN-----LNRSLLTHALLMWSYLFDKLDHFSSETPDSGVSF 100

Db 939 VIQMDTTGRTILENNSGRSNPNFNKEELTAILKFGAEDLFKELE-----GEES 988

Qy 101 EQSIMDGVIIHFSSILSSKGGEENEVKLLLEAKHAQGTYSSTLFGEDHKLSDSES 160

Db 989 EQQEMD-----IDELRLAETRENEVSTSATDELL--SQFKVANPATMEDEEEL-----EER 1038

Qy 161 PNIFWSKLLGKNPMKYPSTPQRNKRQVYFE-----GSEASPTG 203

Db 1039 PHKMDWELI-----PEQRKKVEEERQKELEIYMLPRIRSTKKAQNTDS 1085

Qy 204 DGNNAKRRKASDDVTPDPRVTPDDDERKASGDHMGALSPKVITLQSSCKSSGTDG 263

Db 1086 DSDTESKQQRSSASESETED---SDDDKK-----PKRRGRPRSRVKDLVEG 1130

Qy 264 TLQGN-----DAFLYSGMHSIGIPEDML--ASQDWGKIPDESQRRLHTVLPKMAKLC 316

Db 1131 FTDAEIRRFIKAYKFF-----GLPLERLECLARD--AELVDKSVADL-----KRLG 1174

Qy 317 QVHLSDACTSMVGNFLEVYENHRIYEEBATTFOAFQIALSWIAALLVKQILSHKESLV 376

Db 1175 ELTH--NSCVSAMQYEEQKENASEGKGPQKR--RGPTIKISGV--QNVKSIHQHEEFE 1230

Qy 377 RANSEL-----AFKCSRVEVDYIYILSCMKSLFLEHTQGLQFCFGTNSK 422

```
Db 1231 MLHKSIPVDEEKYKCYLTCRVAAPFVDEWVEDDSRLLLGIYEHGYG--NWELIKTDPE 1289
Qy 423 QSVVSTKLVNESLSGATVRDEKINTKSMRNSSE-----DEECMTKR 464
Db 1290 -----LKLTKILPVET--DKKPGKQLQTRADYLLKLLRKGLEKGAVTGGEAKLKR 1342
Qy 465 -----CSHYSTATRDIEKTIKIKYKQVOKLVQEHKEK 501
Db 1343 KPRVKENKVPRLKEHGBELSPRHSNDPSEGEVXDDGLEKSPMKKQK--KKENKEN 1401
Qy 502 MELLNMYADK-----KOKLETSKVEA-----AVRITCSRTSTQVGD--- 539
Db 1402 EKQSSRKDEGDKERKSKDKKEKPSGDAKSSSKRSQGPVHIITAGSEPVIGDED 1461
Qy 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLDOETFSICKERMVRPVKALQDKDPKGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
Qy 565 -----LEQMHVAKKLADEACWNRINKSWAAKLKVCV 598
Db 1522 DQEHIKLWRNLWIFVSKFTEDFARKLHLKYMAHKRSQEEE-----EQKKDDV 1572
Qy 599 PLOGNNKHF-----SGSSN---ISON 617
Db 1573 ---TGGKKPFRPASGSSRSLISQS 1595

RESULT 12
US-10-732-923-8160
; Sequence 8160, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8160
; LENGTH: 1445
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-732-923-8160

Query Match 5.0%; Score 161; DB 17; Length 1445;
Best Local Similarity 18.0%; Pred. No. 0.0064;
Matches 143; Conservative 113; Mismatches 211; Indels 328; Gaps 35;

Qy 2 ADAPILFGSSLNPSHDVKHVEK-IKIBSCSERTKIFRLYSVCTVEEKALILARQNM-- 58
Db 551 ANVVILFDPFTWNPANDLQADRAYRIGQCRD--VKVFLRLISLGTVEE--INVLRYQYKQL 607
Qy 59 -----NKAVERNLRSLTHALLMWGASYLF-----DKLDH----- 87
Db 608 HCVVVGSENAKRYFEAVQSGKEHGFVGNLFLKRSQGSCLTRDILEREQVGEAGIMT 667
Qy 88 -----PHSETPD-----SGVSFEQSIMGVTHE----- 111
Db 668 ATTWLKGEASAKLETPRDPDQBPNTVBCLYSDMSDEESVGKAVKHVSGPSRTFGSSA 727
Qy 112 -----FSSILSSK-----GGENEVKLCILLAEKHA--QGTYSDDSTL 147
Db 728 QLTLLQCGFKLFEAKCESVQDGNPNVSSDGSDEQPMCLSAKARQAQCKTWDVSCT- 786
Qy 148 FGEDHIKLSDEESNIFWSKLLGKGNPMWKYPDTPQNRKRVQYFEGSEASPKT----- 202
Db 787 --SEH-----QKSDNI-----QTPDEKCVSD---KSEKTLQNVYSSDDDETKCHST 828
Qy 203 -----GDGNAKRVKASDDVTDPP--RV-----TDPVP---DDDER 233
```

```
Db 829 AGHHCWCGDTESESDSVIFPTQDPTQRPVKNPICCKLLLGESDSETEDPVKVNHDDGR 888
Qy 234 KASGK-----DHMGALLESKPVITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDML 289
Db 889 QNSGRNGPVSNNFLYLEN---MTLKSVRKRKGTDDISDESDDIDMF-----PKSRI 936
Qy 289 ASDMG---KIPDESQRRLHTVLKPKMAK--LCOVTLHLSDACTSMVGNFLEVVIENHRIYE 344
Db 937 RKQRATASLKLKSKKENKRNKHYNSPITAKDTNOVCBAGDCSSQ-----VIEDPSSSD 989
Qy 345 EPATTTQAFQIALSWIAALLVKOI-----LSHKESLVRANSELAFKCSRVEVDYIYLSLSC 400
Db 990 DSSS-----VSHVSTKLNHRAETVKQGVQ-----EVAIYH----- 1020
Qy 401 MKSLFILEHTQGLQDFCGTNSKQSVSTVKLVNLSLGSATVRD----- 442
Db 1021 -----SNQNVIGSSRAENHMSRWATRDVFLKQFSQLPANVAVCS 1060
Qy 443 -----EKINTKSMRNSSEDE-----ECMTEKRCSHYST--ATRDIEKTISG 481
Db 1061 SETYKSKVANAVIPTKKQDPPSDGHISPLYISHPVSQKKDVYRTNHTTFFIIGETPRG 1120
Qy 482 IKKK-----YKKQVOKLVQE-----HEEKKM-----ELLN 506
Db 1121 IRRKQFEEMASFYKLSVKFEAEQVTSATSEEQQRLMRDPYSLQHPVKEFFEMDSASELMK 1180
Qy 507 MYADKKQKLETSKSVAAVIRITCSRTSTQVGLDKLLDHNHYERKFDKSEKNECLKSLE 566
Db 1181 SVHEKEERVNKSKGESLLKERPSNDSTL-----SCNDSTN 1217
Qy 567 QMHEVAKKLADEEA 581
Db 1218 KMSQVYNKICEGKS 1232

RESULT 13
US-10-732-923-8335
; Sequence 8335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8335
; LENGTH: 1738
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-8335

Query Match 5.0%; Score 160; DB 17; Length 1738;
Best Local Similarity 17.3%; Pred. No. 0.0099;
Matches 132; Conservative 104; Mismatches 241; Indels 286; Gaps 28;

Qy 2 ADAPILFGSSLNPSHDVKHVEKIKIBSCSERTKIFRLYSVCTVEEKALILARQNM----- 56
Db 1041 ADTVIFDSWNPQNDLQAMRAHRIQQQEVVNIYRFPVTSKSVSEEEILERAKRMVLDHL 1100
Qy 57 -----RONKAVERNLRSLTHALLMWGASYLFDKLDHFFHSSETPDSGVSPQS 103
Db 1101 VIQKLNAGELEKRETKGSNFDKNLSAILRGAELFKE----- 1141
Qy 104 IMDGVTHEFSSILSSKGGENEVKLCIL-----LEAKHAQCTYSDDSTLFGEDH 152
Db 1142 -----DKNDEESKRLSMDIDILERAQVQEEKH---TDETHELLGAPK 1184
Qy 153 I-KLSDEESNIFWSKLLGKGNPMWKYPDTPQNRKRVQYFEGSEASPKTGDGNAKVR 211
Db 1185 VANFCAEDDGSFWSRWIKPDSVVTAEALAPAAARNTKSYVDPSPHD-----RTSKRK 1238
```

QY 212 KKASDDVTPRVTPPPVDDDERKASGDHMGALSPKVITLQSSCKSSGDTGTLGNDADF 271
DB 1239 KKGS-----PPEHTSQRKRTEYFVSTPELLSGTSAQVKGWNGYNLPPKDAQ 1288
QY 272 GLX-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLCQV 318
DB 1289 RFYRTVMKFNHNMQACIAEVEGVVE-----AAPAEAQVELFDA-----1328
QY 319 LHLSDACTSMV---GNP-----LEVNIENHRIYEBPATTFO 351
DB 1329 --LIDGCKESVETGNFEPKGPVLDFFGVPVKANELLKRVQGLQLLSKRI1SYNDPI1SQFR 1386
QY 352 AFQIALSWIA-----ALLVKQILSH-----KESLVRANSELAF---KCSRV 391
DB 1387 -----VLSYLPKSNWCKCGNQWIDARLLGILYHGFNGWEKRLDESGLTKKIAPEL 1442
QY 392 DYIYSILSCMSLFLHTQGLQFDCF-----GTNSKQSVSTKLVBNSLSG---ATVRDE 443
DB 1443 QHHTFPLPRAPNLKERATALLELAAAGGKNTNKA1SKRNKVKVKNLQNFKAPARDR 1502
QY 444 K-----INTK-----SMRNSSE-----DEE 458
DB 1503 RGKSGPANVSLSTKDGPRKTQKAEPLVKBEGEMSDGGEVVEQPKQKMEWCEDVLADE 1562
QY 459 CMTEKRCCHYSTATRDI---EKTISGIKKYK---KQVQKLVQEHKEKMKELLNMYADKKQ 513
DB 1563 IKTLGRLOQTTSADLPKXKVLPRYU1EILGRRIDAIVLEHEE-----DLYKQDRM 1616
QY 514 KLETSKSVEAAVIRITCSRTSTQVGDLLDLDHNYERKFD1BKSEKNECLKSLEQMHVAK 573
DB 1617 TMLR-----WNVVSTFNSLGD-----LNQIYSKLK 1643
QY 574 KKLAEADACWINRIKSWAAKLKVCVPIQSGNKNKHFSGSSNISQ 616
DB 1644 QEKEEBEGV-----GPSHLNGSRNFOR 1665

RESULT 14

US-10-369-493-3210
; Sequence 3210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3210
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3210

Query Match 4.9%; Score 158; DB 15; Length 1015;
Best Local Similarity 20.1%; Pred. No. 0.0066;
Matches 129; Conservative 108; Mismatches 254; Indels 150; Gaps 28;

QY 17 DVKHVEKIKIESCSCERTKIPFLYSVCTVEERKALILARQNMKRONKAVENLRSLTHALLMW 76
DB 225 DIEHWROQTDLKKKESLSRL-----QEBA---TRKTEQADQMSLSLNTLN-----269
QY 77 GASYLFDKLDHFSSETPDSCVGSFEQSIMDGV1HEFSSILSSKGGRENEVKLCILLLEAK- 135
DB 270 -----DIMDTAAALGVPIKGGADPHD1GATILEVFDN1LNLY--EDTEQRLVQTEEAKD 321

QY 136 HAQGTYSSTLFGEDHI-KLSDEES-----PNIFWSKLLGGKPNWPKVPSDT 182
DB 322 KAAAKAKESATRSRSTIQKLTSENKLEQORDEQRKADAFYQKL---KKKGWLETFED 378
QY 183 P-----QNRNRKVQYFEGSEASPKTGDGNGAKKR-KKASDDVTPDPV---TDPVPDD 230
DB 379 PSASLALDQKSLQDQDRWESEKNALNKITNLNKS1EENVALKDKMLQMVVERDSLVDL 438
QY 231 DERKASGDHMGALSPKVITLQSSCKSSGDTGTLGNDADFGLYSMGSHISGIPEDML 288
DB 439 QORQE-----GDIKSLNQOVLDKOKLASKA-----SAGADLKDIOQLRL 478
QY 289 ASDWGKIPDESQRRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEVPVNIENHRIYEBPAT 348
DB 479 ENK-----SLEDQOR-----VOAEKEVLQOOLQTSOTKARLEKVEYTLTKNTASQKMDLETORN 530
QY 349 TFOAQFIALSWIAALVKQI-LSHKESL-----VRANSELAFKCSRVVE- 391
DB 531 EWSKAKVLEEEISRLKKEVDLARKNSAAPTIVEIKPEPNIAESITRVEQLMKEQKPOLP 590
QY 392 ----DYIYSILSCMSL-----FLEHTQGLQFDCFGTNSKQSVSTKLVBNSLSGATVRDE 443
DB 591 LTARDFVQGLVAKLQKQGHKOLDASEDVIRECIG-----LLNRVLK-EPVFDE 638
QY 444 KIN-----TKSMRNSSEDEB---CMTEKRCCHYSTATRDIEKTI1SGIKKYKKQVQ 491
DB 639 KQKEMLOTLQNELNEARLKGDREBQSAQLO1SIIDKYEGEKSELEKEV-----KVHKEIK 694
QY 492 KLVQEHKEKMKELLN-----MYADKKQKLETSKSVEAAVIRITCSRTSTQV---G 538
DB 695 KLAQEITEKTS1LANQOBERDMLKESVEEQ1KNLNADHKAAAE1KVKHQNELTLQRLKDG 754
QY 539 DLKLDHNYERKFD1BKSEKNECLKSLEQMHVAKKLAED 579
DB 755 DLK--ETNLLQKLDTLRQQNESERNRLOADYAAAEKAKLTKD 793

RESULT 15

US-10-425-115-329771
; Sequence 329771, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 329771
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1.pcp
US-10-425-115-329771

Query Match 4.9%; Score 157; DB 16; Length 126;
Best Local Similarity 35.4%; Pred. No. 0.0004;
Matches 40; Conservative 18; Mismatches 47; Indels 8; Gaps 3;

QY 283 IPEDMLASQDWGKIPD---ESQR-----RLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEY 335
DB 10 LPMSSTPSADLHLLPETGVENMRTPKSSFHAELKRELSKLTVLKLPDNLFLANQSLLEY 69
QY 336 VIENHRIYEBPATTFOAQFIALSWIAALVKQI-LSHKESLVRANSELAFKCS 387
DB 70 LNNHLVREPR1ILHAFNL1ALCWRAASFUKYT1ELDHRESLALASDGLNTECN 122

Search completed: July 12, 2005, 01:07:26
Job time : 62.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 09:40:15 ; Search time 7459.33 Seconds
(without alignments)
4033.965 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILGSLNPSHDVKH.....SGNNKHFGSSNISONAPDV 621

Scoring table:

BLOSUM62.
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/ROBINSON929/runat_11072005_102152_29171/app_query.fasta_1.2325
-DB=GenEmbl -OFFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPT=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=ROBINSON929 @CNC 1 12596 @runat_11072005_102152_29171 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.btg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3216	100.0	6554 8 AF213627	AF213627 Arabidops
2	3216	100.0	6571 6 AX063583	AX063583 Sequence
3	2980	92.7	10329 6 AX063582	AX063582 Sequence
4	2971	92.4	10329 8 AF213628	AF213628 Arabidops

C	5	2971	92.4	120965	8	AC026875	AC026875 Genomic s
	6	1033	32.1	695	8	ATH552638	AJ552638 Arabidops
	7	521.5	16.2	4027	8	AK068790	AK068790 Oryza sat
	8	494.5	15.4	467	6	AX063609	AX063609 Sequence
	9	472	14.7	515	6	AX063611	AX063611 Sequence
C	10	442	13.7	519	6	AX063607	AX063607 Sequence
C	11	397	12.3	184118	8	AP005621	AP005621 Oryza sat
	12	332.5	10.3	150597	2	AP005005	AP005005 Oryza sat
C	13	332.5	10.3	156603	8	AP004752	AP004752 Oryza sat
	14	166	5.2	5837	6	AR447512	AR447512 Sequence
	15	166	5.2	7764	6	CQ414477	CQ414477 Sequence
	16	166	5.2	7764	6	AX828387	AX828387 Sequence
	17	166	5.2	7764	9	AF006514	AF006514 Homo sapi
	18	162.5	5.1	4769	6	CQ843036	CQ843036 Sequence
	19	162.5	5.1	4769	9	AK125995	AK125995 Homo sapi
	20	160.5	5.0	4190	10	BC046638	BC046638 Mus muscu
	21	160.5	5.0	7318	10	AB026497	AB026497 Mus muscu
	22	160	5.0	3429	10	AB029919	AB029919 Mus muscu
	23	159.5	5.0	7396	6	AX686226	AX686226 Sequence
	24	158.5	4.9	5946	6	CQ730200	CQ730200 Sequence
	25	158.5	4.9	5998	9	CR456526	CR456526 Homo sapi
	26	158.5	4.9	7450	9	AB191263	AB191263 Homo sapi
	27	158	4.9	5061	10	BC076567	BC076567 Mus muscu
	28	158	4.9	5175	6	BD079934	BD079934 Cancer-as
	29	158	4.9	5175	9	HUMCLIP	M97501 Human cypc
	30	157	4.9	110000	2	PFMAL8P1.11	Continuation (12 o
	31	156	4.9	4597	10	RNO237670	CQ2720760 Rattus no
	32	156	4.9	5176	6	CQ2720760	CQ2720760 Sequence
	33	156	4.9	5857	6	AR274942	AR274942 Sequence
	34	156	4.9	5857	6	AX329796	AX329796 Sequence
	35	156	4.9	5857	9	HSRSTIN	X64838 H.sapiens m
	36	156	4.9	5858	6	A21860	A21860 H.sapiens m
	37	155.5	4.8	6965	9	HSU31906	U31906 Homo sapien
	38	155.5	4.8	7636	6	AX933989	AX933989 Sequence
	39	155.5	4.8	7695	6	AX336423	AX336423 Sequence
	40	155.5	4.8	7695	6	AX933991	AX933991 Sequence
	41	155.5	4.8	7695	9	HSU41740	U41740 Human trans
	42	155.5	4.8	7743	6	AX933993	AX933993 Sequence
	43	155.5	4.8	7761	6	AX933995	AX933995 Sequence
	44	155.5	4.8	8063	6	CQ414705	CQ414705 Sequence
	45	155	4.8	6730	5	AY217130	AY217130 Taeniopvq

ALIGNMENTS

RESULT 1
AF213627

LOCUS

AF213627 Arabidopsis thaliana MOM mRNA, complete cds.

DEFINITION

AF213627 Arabidopsis thaliana (thale cress)

ACCESSION

AF213627.1 GI:8132767

VERSION

AF213627.1

KEYWORDS

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

1 (bases 1 to 6554)

Amedeo P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.

Disruption of the plant gene MOM releases transcriptional silencing

of methylated genes

Nature 405 (6783), 203-206 (2000)

JOURNAL

MEDLINE

20279299

PUBMED

10821279

REFERENCE

2 (bases 1 to 6554)

Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and

Paszkowski, J.

Direct Submission

Submitted (09-DEC-1999) Friedrich Miescher Institute, P. O. Box

2543, Basel 4002, Switzerland

JOURNAL

FEATURES

source

Location/Qualifiers

1. .6554

/organism="Arabidopsis thaliana"

Db	2437	GCTGTAGAGAACCTAAACCGCTCTCTCACGACGACCTGCTCATGTGGGGCGCTCATAC	2496
Qy	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
Db	2497	TTATTTGATAAACTGGATCATTTTTCACAGCAGTGAACCTCCAGATTTCAGAGATTTCATTT	2556
Qy	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
Db	2557	GAACAATCTATTATGACGCGGTGATTCATGAATCTCTGCTCCATATCTTCTTCAAGGT	2616
Qy	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140
Db	2617	GGAGAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGCCCAAGCATGCTCAGGGAAC	2676
Qy	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
Db	2677	TACAGCAGTGATTTCTACTCTATTGTGTGAAGACCATATTAAAGTTGTACAGTGAAGAGAT	2736
Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpProSer	180
Db	2737	CCAAATATATTTTGGTCAAGCTGTGTGGGGGGAAGAAATCCTATGTGGAATACCTTCA	2796
Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
Db	2797	GATACTCCCAAGGAATCGAAAAAGAGTTCACTATTTTGGAGGTTCTGAAGCGAGTCCC	2856
Qy	201	LysThrGlyAspGlyGlyAsnAlaLysArgLysArgLysAlaSerAspValThrAsp	220
Db	2857	AAACCTGGCGGATGTTGGAAATCGAAAGACGGAAGAGGCTTCTGATGATGTCATGAT	2916
Qy	221	ProArgValThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHis	240
Db	2917	CCCCGGGTCACTGATCCGCGCATGATGATGATGAAGAAAGGCTCTGGGAAGGATCAC	2976
Qy	241	MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly	260
Db	2977	ATGGGGGCTTTGGAGTCACCAAAAGTCATAACACTCCAGTCATCATGTAAATCTTCGT	3036
Qy	261	ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle	280
Db	3037	ACAGATGGTACATTTGGATGGAATGATGTTTGGCTTGTATTCTATGGCGAGCCATATC	3096
Qy	281	SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer	300
Db	3097	TCGGAAATCCAGAGATATGTTAGTCTAGATTTGGGGGAATAATCCGGATGATCA	3156
Qy	301	GlnArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis	320
Db	3157	CAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGCAAACTTTGCCAAGTTTTCAT	3216
Qy	321	LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis	340
Db	3217	CTTTCAGATGCTTGCACAGCATGTCGGAAATTTTCTCGAATATGTTATTGAAATATCAC	3276
Qy	341	ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle	360
Db	3277	CGAATCTAGAGAGCCAGCCACTACTTTTCAGGCATTCAGATAGCCCTGAGTTGAT	3336
Qy	361	AlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer	380
Db	3337	GCAGCCTTGTGTGTAAGCAAAATCTTAGCCACAAAGAAATCTCTGGTCCGTGCAAAATCT	3396
Qy	381	GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTrpSerIleLeuSerCys	400
Db	3397	GAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGATATATTATTATTCGATATTGCTCTGC	3456
Qy	401	MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn	420
Db	3457	ATGAAGATCTGTTCTCTGGAGCATACACAAGGTTTCAGTTCGATTGCTTGTACTAAT	3516
Qy	421	SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal	440
Db	3517	TCTAAACAGTCAGTGGTTAGCACAAAACTAGTAAATGAAAGTCTCTCAGGGGCTACAGTG	3576

```

/mol_type="mRNA"
/cultivar="Zhuerich"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATEAT1 and phyA"
310..6315
/function="involved in the maintenance of transcriptional
gene silencing"
/codon_start=1
/product="MOM"
/protein_id="AAF73380.1"
/db_xref="GI:8132768"
/translators="MKDEKIGLGTITRYSLAASIPASVEQETPLRRSSRTPTST
KITPASATSERLAPSPASVSKSGIVKNSTPSPSLRNSRNGKTEVLSQSKGSDN
STRKGTSPDIEQRKQSVBSTDKIPIMASRYRALFRGLKESALVDASNEEL
VVVGCRRIPAGNDQVQKDCPPADAGSRKLPVDETSLOKGTDFPLKSVTEKIV
LDASPIVETGQVPSLENLETKLODKTDCSPANAESKTLFVGETSLVEKPYQ
KFQDNTDCLPPANAESKRLPVGETSLEKDTDFPLKSTTETGKWLVSPIVETRDS
VICSPSTNLETKLVSLGLETDIIVLPKRRDRTABIELDACATVANGDDHVMSSDG
VIPSPGCKNDNRPEMCTCKKQKNGDCNRSVCICVQPVESDNTVDMKTEGP
VTSREYENGQIQHKSDDPKFYSSVPEYWPVQLSDVQLEQCYQLTKSKLSLSL
SKIDLGAETLNSVRKTCUHPYMDASLQKLLTLEHEILDVEIKASGKLUHLDK
MLTHIKNGLKAUVFYQATQTPGELLGNILEDFVQRFQPKSYEHIYSSKNSAIN
NFKNSQCVLLLETACQTIKLRADAPILFGSSLNPSHDVKHVEKIKIBSCSRT
KIFRLYSVCTVEKALILARONKONKAVENRSLTHALLMMGASVLPKLDHFRSS
ETPDGVSFEQSTMDGVIEHFSSILSKSGEENEVKLCILLEAKHAQGYSSDSTLFG
EDHILKSDRESPIFWSKLLGKNPMKPSDTPQNRKRVQVFEGSEASPKTGDCGN
AKRKASDADVDPRVTDPPVDDERRKASGKHGALSPKVITLQSSCKSGTDGTL
GNDAPFGLYSMSGSHISGIDPEMLASQDWGKIPDESORLHLTVLKPMAKLCQVHLSD
ACTSMVGNLEYYVIEHRIYVEBPATTFQAFIALSWIAALLVKQILSHKESIVRANSE
LAFKCRINTEVDYIYSLSCNLSFLEHTGLOGDFCTGNSKQSVSTKLWNESLQAT
VDEKINTKMSRNSDEECMTSEKRSYSTATREDIETKISGKKVKKQVOKLVQEH
BEKMELLNMYADKKQLETSKSYEAIVRITCSRTSTQVGDLLKLDHNYERKPDIK
SEKNECLSQEHOVAKVLADEACWNRINIKSWAAKUKVCVPIQSGNNKHFGSGSN
ISQNAEDVQICNNAYEATYADTNCKMKNVQVPEAENTLGTMSGSTQOVHBMVDVR
NDETMVDVSLRQLTKSOSNEHASITVPEILI PADCQBEFAALNHLSEDQNDIRIT
SAASDTSRVRPEQSLENLSASPEFLNREALVTENRRTSHVGPDTNIDLOQ
NREDCSLDOELIDELAMPVQHLASVVEITRGAESDOYGODICPMPSLACKOPDPAAN
TSENELEALPEQASGEVETIDPAASHQGDQVTCPLSSPTGNOPEANEIGNONI
NTSAEPHVGADVESGDVAIDQETMGAGDQACLSFGSSVGTOSDIGALIEGQNVTV
AQLPTDSMEIVTGVSPVQDQADASPMLSPSGNHPDTAVNIEGLDNTSIAVAPHIS
GSDACMEISEPQVERSTFANLPHFEGVHSGAVTALVPSLNNGTQIAVQVPQ
IPEPVPNDPFLHLEKLRRESNPKTFEKKILKAELEKMAVEQKQALQRAQVQSA
AEHNTTETKEDKLVINMKLANAFSLKTDKVPSPGAPGKIQALQRAQVQSA
LRYNIAPQLOASSFPAPALVAPLQLOQSSFPAPAPAPLOQPOASSFPSPSVRPSAIL
LNFVPCMPQPROPLNSNAPTSTVTPATNPGFSPAPHLNSRPSSTSPVATATPTS
SVPPQALTVSAVSIQQQQEQPQSLSSGLQSNNEVVVCLSDDE"

```

ORIGIN

Alignment Scores:					
Pred. No.:	4,09e-251	Length:	6554		
Score:	3216.00	Matches:	621		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	8	Gaps:	0		
US-10-018-929c-3_COPY_650_1270 (1-621) x AF213627 (1-6554)					
Qy	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis	20		
Db	2257	CGAGCTGATCGGTTATCTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	2316		
Qy	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40		
Db	2317	GTAGAGAAGATAAAAAATCGAGTCATGTTCTGAAAGAACTAAGATATTCCTGATTGTCTCA	2376		
Qy	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys	60		
Db	2377	GTAATGACAGTTGAGAGAAAAGCCCTGATTCGTGCTAGGCAAAATATGCGGCAAAATAAA	2436		
Qy	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80		

```

Qy 441 ArgAepGluLysLeuAenThrLysSerMetArgAenSerSerGluAepGluGluCysMet 460
|
|
|
Db 3577 CGTGACCAAAAGATTAATACCAAGTCGATCGCAATAGCTCAGAGGATGAAGATGCATG 3636
|
|
|
Qy 461 ThrGluLysArgCysSerHisThrLysThrAlaThrArgAepLysGluLysThrLysSer 480
|
|
|
Db 3637 ACTGAGAAGAGATGATGACCATATATAGCACACACAAAGAGATATCGAAAAGCATATTAGT 3696
|
|
|
Qy 481 GlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 500
|
|
|
Db 3697 GGCATAAAAAGAAATACAAAGCAAGTGCAAAAGCTTGTAAGAGCATGAGGAAAAG 3756
|
|
|
Qy 501 LysMetGluLysLeuAenMetTyAlaAepLysLysLysLysLysLysLysLysLysLysLys 520
|
|
|
Db 3757 AAAATGGAGCTGTGTTAAATATGATGACACAAAGCAAGCAAGCAACTTGAACATAGTAAAGT 3816
|
|
|
Qy 521 ValGluAlaAlaValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 540
|
|
|
Db 3817 GTGAAGCAGCAGTAATTCGTTATACCTGTTTCACGGACCATGACTCAAGTGGGTGATCTC 3876
|
|
|
Qy 541 LysLeuLeuAepHisAenTyArgLysLysPheAepGluLysLysLysLysLysLysLysLysLys 560
|
|
|
Db 3877 AAATGCTGGATCATATTAAGAAAGATTTGATGAAATCAAAAGTGAAGAAATGAA 3936
|
|
|
Qy 561 CysLeuLysSerLeuGluGluMetHisGluValAlaLysLysLysLysLysLysLysLysLys 580
|
|
|
Db 3937 TGCCTCAAAAGCTGGAGCAATGACGAGCTTGCAAAAGAAAGATTTGGCTGAGGATGAA 3996
|
|
|
Qy 581 AlaCysTrpLysAenArgLysSerTrpAlaLysLysLysLysLysLysLysLysLysLysLys 600
|
|
|
Db 3997 GCCTGTGGATTAATCGGATAAAGAGCTGGCAGCTAAATTAAGATTTGTGTCCCAT 4056
|
|
|
Qy 601 GlnSerGlyAenAenLysHisPheSerGlySerSerAenLysSerGlnAenAlaProAep 620
|
|
|
Db 4057 CAAGTGGCATACACAGCATTTTAGTGGTTTCATCAACATTTCCCAATAGCTCCTGAT 4116
|
|
|
Qy 621 Val 621
|
|
|
Db 4117 GTA 4119

RESULT 2
LOCUS AX063583
DEFINITION Sequence 2 from Patent WO0100801.
ACCESSION AX063583
VERSION AX063583.1 GI:12541307
KEYWORDS Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1
REFERENCE
AUTHORS Habu, Y., mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 2 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source
1. .6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310_..6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
/translaton="MKDKETGLTGRITITRSLAASPASVEQETPLGRSSRCTPTST
KVITPASATRKSERLAPSPASVSKSGGIVKNSTPSSLRNSRKGTEVSLQSSKGSND
SIRKGTSPDIEQRKQSVBESDTKIPKINGARSFALFRGLKESEALVDASNEEEL
VWVGCRRRIIPAGNDVQKTDCCPPADAGSKRLPVDETSLDKGTFDFPLKSVTETKIV

```

ORIGIN

Alignment Scores:

Pred. No.:	4,11e-251	Length:	6571
Score:	3216.00	Matches:	621
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-018-929c-3_COPY_650_1270 (1-621) x AX063583 (1-6571)

Qy 1 ArgAlaAepAlaPheLysLeuPheGlySerSerLeuAenProSerHisAepValLysHis 20
 Db 2257 CGAGCTGATGCTTTATTTCTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAAGCAC 2316
 Qy 21 ValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
 Db 2317 GTAGAGAGAGATAAAATCGAGTCAATGTTCTGAAAGAACTAAGATATTTCCGATTGTACTCA 2376
 Qy 41 ValCysThrValGluGluLysAlaLeuLysLysLysLysLysLysLysLysLysLysLys 60
 Db 2377 GTATGTACAGTTGAGAGAAAAGCCCTGATTTCTGGCTAGGCAAAATATATCGGCAAAATAA 2436
 Qy 61 AlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 2437 GCTGTAGAGAACTAAACCCCTCTCTCACGACGACTGCTCATGTGGGGGGCGTCATAC 2496
 Qy 81 LeuPheAepLysLeuAepHisPheHisSerSerGluThrProAepSerGlyValSerPhe 100
 Db 2497 TTATTTGATAAATCGGATCATTTTTCACAGCAGTGAAACTCCAGATTCAGGAGTTTTCATTT 2556
 Qy 101 GluGlnSerLysMetAspGlyValLysHisGluPheSerSerLysLeuSerLysGly 120
 Db 2557 GAACAATCTATTATGAGCGCGTGATTCATGAATTCCTGTCATCTTTCTTCCAAAGGT 2616
 Qy 121 GlyGluGluAenGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 2617 GGAGAGAAAATGAAGTCAAGCTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAC 2676
 Qy 141 TyrSerSerAepSerThrLeuPheGlyGluAepHisLysLysLeuSerAepGluGluSer 160
 |||||

Db 2677 TACAGCAGTGATTCTACTCTATTGTTGGTGAAGACCATTAAAGTTGTCCAGATGAAGAGAGT 2736

Qy 161 ProAsnIlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTrpProSer 180

Db 2737 CCAAAATATATTTTGGTCAAAAGCTGTTGGGGGGGAAAAATCCTATGTGGAAATACCCCTTCA 2796

Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200

Db 2797 GATACTCCCCAAGGAATCGAAACGAGTTTCAGTATTTTGGGGTCTCTGAAGCGAGTCCC 2856

Qy 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAsp 220

Db 2857 AAAATGGCGATGGTGAATGCAAGACGAAAGAGCTTCTGTATGATGTCACTGAT 2916

Qy 221 ProArgValThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHis 240

Db 2917 CCCCGGGTCACTATCCCGCAGTAGATGATGATCAAAAGAAAGCCCTCTGGGAGGATCAC 2976

Qy 241 MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly 260

Db 2977 ATGGGGCTTTGGAGTCACCAAAAGTCATTAACCTCCAGTCATCATGTAATCTTCTGGT 3036

Qy 261 ThrAspGlyThrLeuAspGlyLysAsnAlaPheGlyLeuTyrSerMetGlySerHisIle 280

Db 3037 ACAGATGGTACATTTGGATGAAATGATGCTTTTGGCTTGTATTCTATGGGCAGCCATATC 3096

Qy 281 SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer 300

Db 3097 TCTGGAAATCCAGAGGATATGTTAGCTAGTCAAGATTTGGGGGAAAAATACCGGATGAATCA 3156

Qy 301 GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis 320

Db 3157 CAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCAAAACTTTCGCAAGTTTGGCAT 3216

Qy 321 LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis 340

Db 3217 CTTTCAGATGCTTGCAACAGCTGGTCGGAATTTTCTCGAATATGTTATTGAAATCAC 3276

Qy 341 ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle 360

Db 3277 CGAATCTACGAAGAGCCAGCCACTACTCTTTTTCAGGCATTCAGATAGCCCTCGAGTTGGATT 3336

Qy 361 AlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer 380

Db 3337 GCAGCTTGTGTGAAGCAAAATCTTAGCCACAAAGATCTCTGGTCCGTGCATATCT 3396

Qy 381 GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys 400

Db 3397 GAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGGATTATATTTATTTCGATATTGCTCTGC 3456

Qy 401 MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn 420

Db 3457 ATGAAGAGTCTGTTCTCGAGCATACACAAGGTTTGCAGTTTCGATTCTTTGGTACTAAT 3516

Qy 421 SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal 440

Db 3517 TCTAAACAGTCAGTGGTTACACAAAACCTAGTAATAAGTCTCTCAGGGGTACAGTG 3576

Qy 441 ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMet 460

Db 3577 CGTGACGAAAAGATTAATACGAAGTCGATCGAATAGCTCAGAGGATGAAGAGTGCATG 3636

Qy 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480

Db 3637 ACTGAGAAGAGATGTAGCCCATTTAGCACACGAACAGAGATATCGAAAAAGACTATTAGT 3696

Qy 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500

Db 3697 GGCATATAAAGAAATACAGAGAGCAAGTGCAAAAGCTTGTACNAGAGCATGAGAAAAG 3756

Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer 520

Db 3757 AAAATGGAGCTGTAAATATGTATGCAGACAAGAAGCAGAAACTTTGAAACTAGTAAAGT 3816

Qy 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540

Db 3817 GTGGAAAGCAGCAGTAATTCGTATTCTGTTTACGGACCCAGTACTCAAGTGGGTGATCTC 3876

Qy 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560

Db 3877 AAATCTCGATCATTAATTATGAAGAAAGTTTGTGAAATCAAAAGTGAGAAAAATGAA 3936

Qy 561 CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGlu 580

Db 3937 TGCCTCAAAAGTCTGGAGCAAAATGCACGAGTTTCAAGAGAAAGTTGGCTGAGGATGAA 3996

Qy 581 AlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValProIle 600

Db 3997 GCCTGTTGGATTAATCGGATAAAGACTGGCGAGCTAAATTAAGATTGTGTCTCCATT 4056

Qy 601 GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp 620

Db 4057 CAAAGTGGCAATAACAAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCTGAT 4116

Qy 621 Val 621

Db 4117 GTA 4119

RESULT 3

AX063582 LOCUS 10329 bp DNA linear PAT 24-JAN-2001

DEFINITION Sequence 1 from Patent WO0100801.

AX063582 ACCESSION

AX063582.1 GI:12541306 VERSION

Arabidopsis thaliana (thale cress) SOURCE

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.

REFERENCE 1

AUTHORS Habu, Y., Mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.

TITLE Gene involved in epigenetic gene silencing

JOURNAL Patent: WO 0100801-A 1 04-JAN-2001; Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft m.b.H. (AT)

FEATURES

source 1..10329 Location/Qualifiers

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

intron 1009..1295

intron 2551..2673

intron 2753..2867

intron 3114..3506

intron 3681..3973

intron 4896..4975

intron 5218..5777

intron 5883..6082

intron 7481..7615

intron 7772..7914

intron 8071..8153

intron 8319..8451

intron 8630..8718

intron 8919..9000

intron 9212..9284

ORIGIN

Alignment Scores:

Pred. No.: 1.2e-231 Length: 10329

Score: 2980.00 Matches: 618

Percent Similarity: 68.51% Conservative: 0

Best Local Similarity: 68.51% Mismatches: 3

Query Match: 92.66% Indels: 284

DB: 6 Gaps: 3

US-10-018-929C-3_COPY_650_1270 (1-621) x AX063582 (1-10329)

```
Qy 1 ArgAlaSepAlaPheLeuPheGlySerSerLeuAenProSerHisAspValLysHis 20
Db 4172 CGAGCTGATGCGTTTATCTTTTGGGAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC 4231
Qy 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
Db 4232 GTAGAGAGATAAAATCGAGTCATGTTCTGAAAGAACTAAGATATTCCGATTTGACTCA 4291
Qy 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAenMetArgGlnAenLys 60
Db 4292 GTATGTACAGTTGAAGAAAAGCCCTGATTTCTGGCTAGGCAAAATATATGCGCAAAATAA 4351
Qy 61 AlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyr 80
Db 4352 GCTGTAGAGAACCTAAACCGCTCTCTCACGCACGCACTGCTCATGTGGGGGGCGTCATAC 4411
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 4412 TTATTTGATAAATCGGATCATTTTTCACAGCAGTGAATCTCCAGATTCCAGGAGTTTCATTT 4471
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 4472 GAACAATCTATTATGGACGGCGTGATTCATGAATCTCGTCCATACCTTTCTCCAAAGGT 4531
Qy 121 GlyGluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 4532 GGAGAAGAAATGAAGTCAAGCTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAC 4591
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 4592 TACAGCAGTGAATCTACTTATTTGGTGAACACCATATTAAGTTGTTCAGATGAAGAGAGT 4651
Qy 161 ProAsnIlePheTyrSerLysLeuLeuGlyLysAsnProMetTyrLysTyrProSer 180
Db 4652 CCAATATATTTTGGTCAAGCTGTGGGGGGAAGAAATCTCTATGTGGAATACCTCTCA 4711
Qy 181 AspThrProGlnArgAenArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 4712 GATACTCCCCAAAGGAATCGAAACGAGTTTCAGTATTTTGGAGGTTCCTGAAGCGAGTCCC 4771
Qy 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp 220
Db 4772 AAACTCGCGATGTTGGTGAATGCAAGAGGAAAGGAGGCTTCTGATGATGTCACTGAT 4831
Qy 221 ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 4832 CCCCAGGTCACTGATCCCGCAGTAGATGATGAAGAAAGGCTCTGGGAGAGATCAC 4891
Qy 241 Met----- 241
Db 4892 ATGGGTAAATPAGTTTAATTTCTGCTCCGACCTCTAGTGTTCATTGATTATGCAACTA 4951
Qy 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 4952 CTTTGTGCTACTATCTTCTACAGGGCT-TTGGAGTCACCAAAAGTCATAACACTCCAG 5010
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAenAspAlaPheGlyLeu 273
Db 5011 TCATCATGTAAATCTTCTGGTACAGATGGTACATTTGATGGAATGATGCTTTTGGCTTG 5070
Qy 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTyr 293
Db 5071 TATTTATGGGACCCATATCTCTGGAATCCAGAGGATATGTTAGTGTCAAGATTGG 5130
Qy 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
Db 5131 GGGAAAAATACCGATGAATCAACAGAGGAGGCTCCACACTGTTTAAAGCCGGAAGATGCA 5190
Qy 314 LysLeuCysGlnValLeuHisLeuSer----- 322
Db 5191 AAATCTTGGCCAAAGTTTGTGCATCTTTC-AGTAAAGTGGCGCTTTTTCACCTCCCAACTTATT 5249
```

```
Qy 322 ----- 322
Db 5250 TTAGCCTTGCATATGCTTATATATATATAGCTGTAAGTCAACTGTAGTTGTTACCTGATTTCCGTG 5309
Qy 322 ----- 322
Db 5310 TTACAGCCAAATGTGAGAGTTTATTTCTTCACTATATATCCATCCGTTTAAGCATATTTTA 5369
Qy 322 ----- 322
Db 5370 TTTCTTATATCTGCTTTCGTTTACCCAATGCACCTGTTAAAAATGAGCAACTGCTGCACAAAAC 5429
Qy 322 ----- 322
Db 5430 AGTAGTAGTTATATGTGCTCATGTTCATTTGTTTATTGAAGCAAAAGAAATTTCTGTCT 5489
Qy 322 ----- 322
Db 5490 ACTTTACATGATCCATCTGTGGGAGTATATAACTATATATAACCTTAGGCCCTTTGTACCT 5549
Qy 322 ----- 322
Db 5550 GGCTGATCAAAGACATGTCAAAAGCTTTATCTGTTTCGCTGTTGGTATAGAACTAATACAG 5609
Qy 322 ----- 322
Db 5610 TGTCTGATGCTATTTTAAAGGTAGTCTTATGCTTTCACATATTTGGCTAATAGATGTTTCCG 5669
Qy 322 ----- 322
Db 5670 CTGTCGTGTCATATATCTTCTGTGATTTATCACGGTGTCCGCTCTATCAAAATTTGTACTAA 5729
Qy 323 -----AspAlaCysThr 326
Db 5730 AAGGTATTTTGAATGTGATGTTTAAACAGATTTATTTGTTTTCAGGATGCTTGACACA 5789
Qy 327 SerMetValGlyAenPheLeuGluTyrValIleGluAenHisArgIleTyrGluGluPro 346
Db 5790 AGCATGTCGGAATTTTCTCGAATATGTTTATGAAATCACCAGNAATCTACGAGAGCCA 5849
Qy 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
Db 5850 GCCACTACTTTTCAGGCATTTCCAGATAGCCCT-GGTATGACAGCATTTTACTTTTGATTAAT 5908
Qy 357 ----- 357
Db 5909 TATGCATTTGTTTCTTTCATCATCTGCTTGTGTTTAGAATGTCTCTCAGAGGAGCAGCACTCC 5968
Qy 357 ----- 357
Db 5969 TTTAGTTTAACTTTCCAATCATAGGATTCAAATATCCATTAACGTGCGCTTTTGATCGCTG 6028
Qy 358 -----SerTrp 359
Db 6029 CATATAATATGAATPAGTTGACATACCTGAATACGTTGTTTAATAATGCAATTTTCAGAGTGG 6088
Qy 360 IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAen 379
Db 6089 ATTGCAGCCTTGTGTTGGTAAAGCAATTTCTTAGCCACAAGNAATCTCTGGTCCGTGCAAAAT 6148
Qy 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
Db 6149 TCTGAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGATATATTTATTCGATATTGTCC 6208
Qy 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 6209 TGCATGAAGAGTCTGTTCTCTGGAGCATACAAAGGTTTGCAATTCGATGCTTTGGTACT 6268
Qy 420 AsnSerLysGlnSerValValSerThrLysLeuValAenGluSerLeuSerGlyValThr 439
Db 6269 AATTTCTAAACAGTCAGTGGTTAGCACAAAACACTAGTAATGAAAGTCTCTCAGGGGCTACA 6328
```

Qy	440	ValArgAspGluLysValIleAenThrLysSerMetArgAsnSerSerGluAspGluGluCys	459
Db	6329	GTGCGTGCAGCAAGATTATACGAAGTCGATCGGAATAGCTCAGAGGATGAAGAGTGC	6388
Qy	460	MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle	479
Db	6389	ATGACTGAGAAGAGATGATGCCATTATAGCACAGCAACAAGAGATATCGAAAGACTATT	6448
Qy	480	SerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu	499
Db	6449	AGTGGCATAAAGAAAGAAATCAAGAAGCAAGTGCAGAAAGCTTGTAAGAGCATGAGGAA	6508
Qy	500	LysLysMetGluLeuLeuAenMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys	519
Db	6509	AAGAAATGAGCTGTTAAATATGATGTGCAGACAAGAGCAGAACTTGAACCTAGTAAA	6568
Qy	520	SerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539
Db	6569	AGTGTGGAACGACAGTAATTCGTATTACCTGTTTCCGAGCAGTACTCAAGTGGGTGAT	6628
Qy	540	LeuLysLeuLeuAspHisAenTyrGluArgLysPheAspGluLysSerGluLysAen	559
Db	6629	CTCAAACTGCTGGATCATATATGAAAGAAAGTTTGATGAAATCAAAAGTGAAGAAAT	6688
Qy	560	GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLeuAlaGluAsp	579
Db	6689	GAATGCTCAAAAGCTCGAGCAAAATGACAGAGTTTCAAGAGAAAGTTGGCTGAGGAT	6748
Qy	580	GluAlaCysTrpIleAenArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro	599
Db	6749	GAAGCTGTTGGATTATCGGATAAAGAGCTGGGACACATAAATTAAGTTTGTGTTCCC	6808
Qy	600	IleGlnSerGlyAenAenLysHisPheSerGlySerSerAenIleSerGlnAenAlaPro	619
Db	6809	ATTCAAGTGGCAATAACAAGCATTTTAGTGGTTTCATCAACATTTTCCCAAAATGCTCCT	6868
Qy	620	AspVal	621
Db	6869	GATGTA	6874

RESULT 4
AF213628
LOCUS AF213628 10329 bp DNA linear PLN 01-JUN-2000
DEFINITION Arabidopsis thaliana MOM gene, complete cds.
ACCESSION AF213628
VERSION AF213628.1 GI:8132769
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 10329)
AUTHORS Amedeo,P., Habu,Y., Afear,K., Scheid,O.M. and Paszkowski,J.
TITLE Disruption of the plant gene MOM releases transcriptional silencing of methylated genes
JOURNAL Nature 405 (6783), 203-206 (2000)
MEDLINE 20279299
PUBMED 10821279
REFERENCE 2 (bases 1 to 10329)
AUTHORS Amedeo,P., Habu,Y., Afear,K., Mittelsten Scheid,O. and Paszkowski,J.
Direct Submission
TITLE Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland
JOURNAL
FEATURES
source Location/Qualifiers
1..10329
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATEAT1 and phyA"

mRNA

CDS

```
join(705..11008,1296..2550,2674..2752,2868..3113,
3507..3680,3974..4895,4976..5217,5778..5882,6083..7480,
7616..7771,7915..8070,8154..8318,8452..8629,8719..8918,
9001..9211,9285..10047)
/product="MOM"
join(1301..2550,2674..2752,2868..3113,3507..3680,
3974..4895,4976..5217,5778..5882,6083..7480,7616..7771,
7915..8070,8154..8318,8452..8629,8719..8918,9001..9211,
9285..9808)
/function="involved in the maintenance of transcriptional
gene silencing"
/codon_start=1
/product="MOM"
/protein_id="AAF327381.1"
/db_xref="GI:8132770"
/translation="MKDKBKIGLGRITTYRSLAASIPASVFEQETPLRRSRSGTST
KITTPASATRKSRAPSPASVSKSGGIVRNSPTSSLRNRKSKTEVLVSQSGSDN
SRKGTSPDIQRKDSVEESTDKIPIMARSYRALFRGLKESKALVDAGPNEKIV
VVGCSRRIPAGNDVQGTDCPPADAGSKELPVDETSLDKGTLPFLKSVTEKIV
LDASPIVETDDSVIGSPSENLETKLODKGTDCSPANAESKTLPGVSTLEKEYPQ
KFQDDNTDCLPANAESKELPVGETSLEKTDFFPLKSTTGTGKMLYASPIVETRDS
VICSPSTNLETQKLIVSKTGLTETDIVLPLKRRDTRAEIELDACATVANGDDHMSDDG
VTPSPGCKNDRNPEMCKTKRKQKVGQNRVCSCIVQPVEESNDVQKMETGPG
VTSRYEENGQIQHGKSDPKFYVYVWVQVLSQVLEQYQCTLFSSKLSLSS
SKIDLCAETLNSVRKTCDFPVMDASLKQLLTKNLEHEILDVDEIKASGKLHLJDK
MLTHIKNGLKAVVEYQATQTPGELLGNILEDFVQGRFGPKSYEHGIYSSKNSAIN
NFKESQCCVLLLETPACSTQIKLLRADAPILFGSLNPSHDVHKVEKIKIESCSERT
KIFRLYSCTVEEKALILARQNKQKAVENLRSLTHALLMWGASYLFDKLDHFHSS
ETPDGVSPEIMDGIHFESSILSSKGBENEVKLCULLAKHAQGTYSYSDTLFG
EHIKILDSBPENIIFWSKLGGKNPMWYPSDTPQNRKRVQYFEGSEASPKTDGDN
AKRKKAASDDVDPRTVDPVDDDERKASGKDHMGALSPKVTILQSSCKSGDTGL
DGNDAFLYSGMSHISGIPEDMASODGKI PDESORRLHTVLKPMWAKLCOVHLSD
ACTSMVGNFLEYVIEHRIYERPAITFQAFQIALSHIALLVKQLSHKESLVRANSE
LAFKCRVEDIYIISLSCMSLFLHETQLOFDFCGTNSKQSVYNNKTLNVSLSGAT
VDEKINTSMRNSDEDECKRCSHYSTATRDIKTISIKYKVKQKVLQVQEH
SEKNECLSLQEKHDAKLAEDACNIRIKSWAAKLKVCVIOGNNKHFSGSN
EKKMELNLYADKKKLETSKSEVAIVITCSRTSQVGLKLIDHNYERKFDRIK
ISONA PDVOI CNNANVEATYADTNCMASKVNQVPEAENTLGTMSGGSTQVHEMDVR
NDETHVSALREQLTKSQSNEHASITVPEILLI PADCOSEFAALNVHLSEDQNCDRIT
SAASDDVSRRVPSVQSLNLSASFSLNREALVTTENRRTSRSHVGTDTNILDQOQ
NREDCSLDQEI PELAMPVOHLASVVETRGAAESDQYQDQICMPSPSLAGKQDPAAAN
TRESNLEEAIEFQESAGSETVETTFDAASHQDQVTCPLSSPTGNQPAPEANIEGNI
NTSAEPHVAGPDVSGDYAVIDQETMGQAQACSLPSGSGVTQSDIGANI EGNQVTV
AOLPTDGSDAVTVGSPVSDCAODASPMPLSPGNHPDPTAVNIEGLDNTAVHHS
GSDACAMEISEPGQOVERSTFANLHEGGVEHSAGVTALVPSLLNNGTQIAVQVPQ
IPFPVNDPFLHELEKLRRESNSKTFPEKKSILKAELEKMAVEQAERFRKHEVE
ABHNTRTTKIEKDKNLVIMNKULANAFLSKTDKKVSPSGARGKIQQLAQAQVSA
LRNYTAPQQLQASFPAPALVAPLQLOQSSFPAPGAPLQFQSPSPSSVRPSALL
LNFVCPMPQPRQPLISNTIAPTSPVTPATNGLRSPAPHLNSYRPSSTPVTATPTS
SYPPQALTYSAVSIQQQEQPQQQLSSLSQSNNEVVCLSDDE"
```

ORIGIN

Alignment Scores:
Pred. No.: 6,45e-231 Length: 10329
Score: 2971.00 Matches: 616
Percent Similarity: 68.40% Conservative: 1
Best Local Similarity: 68.29% Mismatches: 4
Query Match: 92.38% Indels: 284
DB: 8 Gaps: 3

US-10-018-929C-3_COPY_650_1270 (1-621) x AF213628 (1-10329)

Qy	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisepValIleHis	20
Db	4172	CGAGCTGATGCGTTTATCTTTTGGAAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	4231
Qy	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40
Db	4232	GTAGAGAGAGTAATAAATCGAGTCATGTTCTGAAAGAACTAAGATATTCGATGTACTCA	4291
Qy	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAenMetArgGlnAenLys	60
Db	4292	GTATGTACAGTTGAAGAAAAAGCCCTGATTCCTGGCTAGGCAAAATAAGCGGCAAAATAA	4351

Qy 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 4352 GCTGTAGAGAACCTAAACCGCTCTCTCAGCAGCAGCTGCTCATGTGGGGCGCTCATAC 4411
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 4412 TTATTTGATAAATCGGATCATTTTCACAGCAGTGAACCTCCAGATTACAGGAGTTTCATTT 4471
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 4472 GAACAATCTATATGACGGCGTGATTCTGAATTTCTCGTCAATCTTTCTTCCAAAGGT 4531
Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 4532 GGAGAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAT 4591
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 4592 TACAGCAGTGATTTCTACTTATTTGGTGAACCCATATTAAGTTGTTCAGATGAAGAGAT 4651
Qy 161 ProAsnIlePheTrpSerLysLeuGlyLysAsnProMetTrpLysTyProSer 180
Db 4652 CCAATATATTTTGTCAAGCTGTTGGGGGAAAAATCTATGTGGAATACCTTTCA 4711
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 4712 GATATCTCCCAAGGAATCGAAAAAGAGTTCAGTATTTTGGAGGTTCTGGAAGCGAGTCCC 4771
Qy 201 LysThrGlyAspGlyGlyAsnAlaLysValArgLysValAlaSerAspAspValThrAsp 220
Db 4772 AAACTGCGATGGTGGAAATGCAAAAGCGAAGAGGCTCTCGATGATGTCACTGAT 4831
Qy 221 ProArgValThrAspProProValAspAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 4832 CCCCAGGTCACTGATCCGCGAGTAGATGATGATGAAGAAAGGCTCTGCGGAAGATCAC 4891
Qy 241 Met----- 241
Db 4892 ATGGGTAAATAGTTTAATTTCTGCTCGACCTCTAGTGTTCATTGATTATGCAACTA 4951
Qy 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 4952 CTTTGCTGACTATCTTTCTCAGCGGCT-TTGAGATCACCAAAAGTCATAAAGCTCCAG 5010
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu 273
Db 5011 TCATCATGTAATCTCTGTTACAGATGGTACATTTGGATGGAATGATGCTTTTGGCTTG 5070
Qy 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGluAspTrp 293
Db 5071 TATTTCTATGGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGG 5130
Qy 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
Db 5131 GGGAAAATACCGGATGAATCAAGAGGAGGCTCCACACTGTTTTTAAAGCCGAAGATGCA 5190
Qy 314 LysLeuCysGlnValLeuHisLeuSer----- 322
Db 5191 AAACCTTGCCAAAGTTTTCATCTTTC-AGTAAGTGGCCTTTTTCACCTCCACAATTTATT 5249
Qy 322 ----- 322
Db 5250 TTAGCCTTGATATGCTTATATAGCTGATTGCAACTGATGTTTACCTGATTTCTCTG 5309
Qy 322 ----- 322
Db 5310 TTACAGCAAAATGTGAGAGTTTATTCTTCAACTATATCCATCCGTTTAAAGCATATTTTA 5369
Qy 322 ----- 322
Db 5370 TTTCTTATATCTGGCTTGTACCAATGCATCTGTTAAATGAGCAACTGCTGCACAAAC 5429

Qy 322 ----- 322
Db 5430 AGTAGGTAGTTATGTGCTCATGTCTCATTTGTTTATGAAGCAAAAGAAATTTCTGTCT 5489
Qy 322 ----- 322
Db 5490 ACTTTACATGATCCATCTGTGGGAGTATATAACTATATATAACCTTAGGCCCTTTGTACT 5549
Qy 322 ----- 322
Db 5550 GGCTGATCAAAAGACATGTCAAAAGTTTATCTGTTCTGTTGGTATAGAAACTAATACAG 5609
Qy 322 ----- 322
Db 5610 TGTCTGATGCTATTTTAAAGTAGTCTTATGCTTTCACATATTTGGCTAATAGATGTTCCG 5669
Qy 322 ----- 322
Db 5670 CTGTCGTGCCATATATCTTCTGTGATTTATCAGGGTGCTCCGCTCTATCAAAATTTGACTAA 5729
Qy 323 -----AspAlaCysThr 326
Db 5730 AAGGTATTTTGCATATGTGATTGGTTAAACAGATTATTTTGTTCAGGATGCTTGACACA 5789
Qy 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
Db 5790 AGCATGTCGGAATTTTCTCGAATATGTTTATTGAAATCAACCGAATCTACGAAGAGCCA 5849
Qy 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
Db 5850 GCCACTACTTTTCAGGCATTTCCAGATAGCCCT-GGTATGACAGCATTTTACTTTTGATAATT 5908
Qy 357 ----- 357
Db 5909 TATGATTTGTTCTTCTCATCTGCTTTGTTTGAATGTCCTCAGAAAGCAGCAGCTCC 5968
Qy 357 ----- 357
Db 5969 TTTAGTTTAACTTTTCCAATCATAGGATTCAAATATCCATTAACCTGCGCTTTGATCGCTG 6028
Qy 358 -----SerTrp 359
Db 6029 CATTAATATGAATAGTTGACATACTGAATACGTTGTTTAATAGCATTTTCAGAGTTGG 6088
Qy 360 IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
Db 6089 ATTGACGCTTGTGTTGTAAGCAAAATCTTAGCCACCAAGAAATCTCTGGTCCGTGCAAT 6148
Qy 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
Db 6149 TCTGAATTAGCTTTCAATGCTCTAGAGTAGAGTGGATTATATTTATTCGATATTTGTC 6208
Qy 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 6209 TGCATGAAGAGTCTGTTCTTGGACATACCAAGTTTGCAAGTTTCGATTTGTTTGGTACT 6268
Qy 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyValaThr 439
Db 6269 AATTCTAAACAGTCAGTGGTTAGCACAAAAGTAAAGTAAAGTCTCTCAGGGGCTACA 6328
Qy 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
Db 6329 GTGCGTCAACAAAAGATTAATACGAAGTCGATGCGAATAGCTCAGAGGATGAAGAGTGC 6388
Qy 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
Db 6389 ATGACTGAGAAGAGATGATGACCATTTATAGCACCAACAGAGATATCGAAAAGACTATT 6448
Qy 480 SerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
Db 6449 AGTGGCATAAAAAGAAATACAGAAGCAAGTGCANAAGCTTTGTACAAAGCATGAGGAA 6508
Qy 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519

Db	6509	AAGAAAATGAGCTGTTAAATATGATATGCAGACAAGAGCAGAACTTGAACCTAGTAAA	6568	
Qy	520	SerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539	
Db	6569	AGTGTGGAACGACGAGTAATTCGTATTACCTGTTACCGGACCACTACTCAAGTGGGTGAT	6628	
Qy	540	LeuIysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn	559	
Db	6629	CTCAAACTGCTGGATCATATTAATGAAAGAAAGTTTGATGAATCAAAAGTGAGAAAT	6688	
Qy	560	GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLeuAlaGluAsp	579	
Db	6689	GAATGCCTCAAAAGCTGCGAGCAATGCACGATGTTCCAAAGAGAAAGTTGGCTGAGGAT	6748	
Qy	580	GluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCysValPro	599	
Db	6749	GAAGCCTGTTGGATTATCGATAAAGAGCTGGCGAGCTAAATTAATAAGTTGTGTCC	6808	
Qy	600	IleGlnSerGlyAsnLeuHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro	619	
Db	6809	ATTCAAAAGTGGCAATAACAAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCT	6868	
Qy	620	AspVal 621		
Db	6869	GATGTA 6874		
RESULT 5				
AC026875/c				
LOCUS				
DEFINITION	AC026875	120965 bp	DNA	linear
				PLN 28-JUN-2000
				Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
				I, complete sequence.
ACCESSION	AC026875			
VERSION	AC026875.4	GI:7651711		
KEYWORDS	HTG.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.			
REFERENCE	1	(bases 1 to 120965)		
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,			
	Khan, S., Kim, C., Altafi, H., Bei, Q., Chin, C., Chiou, J., Choi, E.,			
	Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T.,			
	Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,			
	Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,			
	Southwick, A., Thaveri, A., Toriumi, M., Vayaberg, M., Yu, G.,			
	Federpsiel, N.A., Theologis, A. and Ecker, J.R.			
TITLE	Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome			
JOURNAL	I			
REFERENCE	2	(bases 1 to 120965)		
AUTHORS	Ecker, J.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-MAR-2000) Arabidopsis thaliana Genome Center,			
	Department of Biology, University of Pennsylvania, 38th Street and			
	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
REFERENCE	3	(bases 1 to 120965)		
AUTHORS	Ecker, J.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,			
	Department of Biology, University of Pennsylvania, 38th Street and			
	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
REFERENCE	4	(bases 1 to 120965)		
AUTHORS	Chen, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q.,			
	Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C.,			
	Chiou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N.,			
	Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,			
	Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,			
	Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vayaberg, M.,			
	Yu, G., Davis, R., Federpsiel, N., Theologis, A. and Ecker, J.			
TITLE	Direct Submission			
JOURNAL	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,			

COMMENT	Department of Biology, University of Pennsylvania, 38th and			
FEATURES	Hamilton Walk, Philadelphia, PA 19104-6018, USA			
source	On Apr 27, 2000 this sequence version replaced gi:7543638.			
	Location/Qualifiers			
	1. 120965			
	/organism="Arabidopsis thaliana"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:3702"			
	/chromosome="1"			
	/clone="T6D22"			
	join(410..559,596..758,828..865,1681..1731,1903..2253,			
	2562..2999,3096..3146,3240..3322,3493..3659,3781..3819,			
	3992..4021,4069..4246,4597..4703,4824..4952,5058..5122,			
	5239..5394,5549..5709,5812..5843,6103..6266,6382..6461,			
	658..6723,6881..7170,7235..7264,7463..7570,7663..7760,			
	7939..7999,8364..8440,8552..8623,8702..8778,8898..8978,			
	9337..9439,9593..9751,9883..10281)			
	/notes="similar to AT.I.24-9 gene product gb AA07881.1;			
	similar to EST emb Z35021.1"			
	/codon_start=1			
	/evidence=not experimental			
	/product="T6D22.1"			
	/protein_id="AAF79821.1"			
	/db_xref="GI:8778822"			
	/translation="MICHNFQGRNRLDLSLPTKTNLSQTIIVAPKISFYVNRCS			
	TVTKIKVTGKSPSSPNOADEKCGKRYVAASRGVCGVLLRLNQLNQLKQKPRD			
	LTYNKILTCGIDGGPMLVEFGFESDFLASKFIDFLLRWSPPHNTSTNSVVPV			
	SQVSDSARRRNLVAASISLYFYVTRTKLSHVLMSLIPHTTHFAFSINHLTSRTS			
	SLRSSPPTFLARSICFNSSIRFLSDMPKQKQKHABQKQVAPKQMDAPRESGDS			
	SATVAEAVNNQFGSLKSNSTNAPVLSQTSNHRVNLVWPKSYSTKLSQSSSAT			
	EVGTSAVSQIGSGDTVGLNLSKIFGGLNLEKFSVDKSTYCHQIRATTFYKPFNE			
	KTDQETRTMIEMVSKLATLEVLSKHSGLFMYAGHGAGAYAKNSFGNIFIVSYAQG			
	VILLADDLVDYDMDKIPHITVVQIYFKKHDFEANAIFYSPFLWRLGEILRLQK			
	QNHGQRPDLDDYDVKVQLKGNVDLHVAVVTAVELGNGKQPFYSTSEIISFCRK			
	WRLPTNHWLFDSTKRSVTFAPADFALCEBEGIATSVCRALDEVADISVPAKSKHVQV			
	GEILEGLVARIVSSQSSRDENVLRDHPPPCCDGNLDLGLSREICAAHRSNEKQOM			
	RALLRSVGFSCFSDVEWFGDESHPSADKSVITKFLQSPADYSTKLQEMVWLKKE			
	KRLPAAPKYNPHRAEDISPONLFYKLVVHSDSGFRYHKEMHMFSLWPLYRGF			
	FVDINLFKNGKRDLMALKSIDNASENDGRGEKGLADGLADMLKMWFLYKLRFL			
	IRNGLSILFKDGNAAKYTYLQMKITGTSQKQKELCKWLDEWAAVIRKCCNDOLS			
	STYLSAEAPFLEQYAKRSRPHILIGSAGNLVTRTEDFLAIVDGDDEGLVKKQGV			
	TPATPEPAVKEAVQDEGLIVFPFVSIELMDGSGIPGSALCSALCEKMDVKGDD			
	RPVHTLMDGLVKGYKWPVADERKPKQSIMLADKNAPNEDVNRQYLEDRCRTRAGAV			
	PVIAVSEGTDTNPYSDALAVFMFRVQVHNGPKLDKSSNAGVLLMFLFYHGYGN			
	RNEFSELTERFGSLKMLPKLSDRTPLDPVKSVLEEGIDLEFNLSRRHRLSTKNG			
	TYAAETWKEQLDRTLVANSEYLSIQIPFESMVHQVREELKTIAGDKYKPSSEKR			
	HTAETVFAINPATQVLSLEKLAANPTMRSFLEGGKKSIOEKLESHVTLAHKRS			
	HGVTAVASQHLNREVVELTELIYNDMAALTAHVGSVDGTVTVSKNENPHTLMT			
	AEGVTAKENTLPQLYLEGKASRLVIDPPVVISGPLEFF"			
	join(11454..11915,12009..12885,12925..13073,14200..14235,			
	14754..14777,14903..15370,15466..16353)			
	/note="putative elongation factor 1-a			
	gb AAFI3128.1 AF12480.1; similar to ESTs gb AW398653.1,			
	gb Z17990.1, gb AW220334.1, gb AW622026.1,			
	gb J0439770.1, gb T45275.1, and gb AW267999.1,			
	gb T04414.1, gb AW221780.1"			
	/codon_start=1			
	/evidence=not experimental			
	/product="T6D22.2"			
	/protein_id="AAF79822.1"			
	/db_xref="GI:8778823"			
	/translation="MGKEKPHINIVIGHVDSGKSTTGTGHIYKLGIDKRVIERFEK			
	EAEKMRKSPKAWLKLKABEREGITIDALWKFTTKYCTVTDAPGRDFIKNM			
	ITGTSQADCAVLIIDSTTGGFEAGISKDQTSREHALLAFTLGKQIMCCCKMDATTP			
	KYSKARYDEITKEVSYLKKGYNPKIPVPVISGEGEDNMIERSTNLDWYKPTLLE			
	ALDQINPEKPSDKPLRLQDVQYKIGGTGTPVGRVETGMKPGWVTFAPTLTTE			
	MKSVEWHESLEALPCDNGVFNVAVKDLRGVANSKDEIEKEPKFKNGDAGVMVM			
	VNGHGOINGYAPVLDCHTSHIAVKESEILTKIDRRSGKEIEKEPKFKNGDAGVMVM			
	TPTPKVPWFTESEYPLGRPAVRDMQTAVGVIKSVDDKQPTGAKTVKAAVKGSK			
	DVSFTITVWLVACLSCEFSPSPSRCSNWLDRWRIYTHILPHHPSPCLDVIYC			
	MIHHNFVLTWKEKPHINIVIGHVDSGKSTTGTGHIYKLGIDKRVIERFEKAAE			
	MNKRSPKAWLKLKABEREGITIDALWKFTTKYCTVTDAPGRDFTIKNMITGT			
	SQADCAVLIIDSTTGGFEAGISKDQTSREHALLAFTLGKQIMCCCKMDATTPKYSK			

ARYDEIITKEYSVYLKVGYNPDKIPFVPISGFEGDNMIERSTNLDWKVGTLLLEALDQ
INPKRSDPLRLPLQDVYKIGIGITVPVGRVETGMKPGMVVTRAPTGLTIEVKSV
EMHESLEALPDGVNFVNAVKOLKRGYVNASNKDDPAKGAANFTSQVILIMHP
GQINGVAPLDCHTSHIAVKFSEILTKIDRRSGKEIEKEPKFLKNGDAGWKMVTPK
PMVTESEYPLGRFAVRDMQTVAVGVKISVDKDDPTGAKVTAKAAVKGAK"
complement (join 19245. .20132,20226. .20687))
/notes="putative elongation factor 1-a
gb|AAFI3128.1|AF124801.1; similar to ESTs gb|AW034714.1,
gb|A1812185.1, emb|Z17990.1, gb|AW267999.1, gb|N96009.1,
dbj|JAV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"
/translation="MGKEKPHINIVIGHVDSGKSTTGHILYKLGIDIKRVIREFK
EAAEMKRSPKAWLMDLKAERERGITIDIALMKFTTYYCTVIDAPGRDPIKMM
ITGTSQADCAVLIDSTTGGFEAGISKDGOTREHALLAFTLVGVKQMICCCNKM DATTP
KYSKARDEIIEKVSLLKGVYNPDKIPEVPSGFEEDNMIERSTNLDWKVGTLLLE
ALQDINPKRSPKPLRLPLQDVYKIGIGITVPVGRVETGMKPGMVVTRAPTGLTIE
VKSVMHESLEALPDGVNFVNAVKOLKRGYVNASNKDDPAKGAANFTSQVIL
MNFHQINGVAPLDCHTSHIAVKFSEILTKIDRRSGKEIEKEPKFLKNGDAGWKM
TPTKPMVTESEYPLGRFAVRDMQTVAVGVKISVDKDDPTGAKVTAKAAVKGAK"
complement (20787. .20857)
/product="tRNA-Lys"
/notes="codon recognized: AAA; putative tRNA-Lys"
/evidence=not experimental
complement (join 21956. .22189,22535. .22600,22713. .22929,
23041. .23126))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translation="MCLNRLPSATPIPRAREERECOSMFCFRNHQILLSSFLCKRPA
IIMKGGSGSGSGGSPPTAAALQKQKALLQREVDTISVDNFTQIVNRSV
SDPPVNSQETIYMMEKARWQAASLLKLVSSELKQTALIFSGFASLNDHVEQRIEF
DQAEKTNLLARADASANKLESHYSSAQRLLTI"
join (23576. .23689,24080. .24258,24347. .24522,24635. .24726)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translation="MTLGARLVAPMIILLFPIELVKAETVITLPTFTSDKLALSNV
KAQLONGICSWCSFYSLSDTTCLLSPCLLLVQIKEDTAMFVKCVPWCKKCLKLG
NLWEDLGKMEGDEIEVGVDCVDSRACTKVEIHSYPTFMIFYNGEVSCKYKGRKD
VESLKAFVVEETEKAQAQLEDKEL"
complement (join 25423. .25563,25675. .26048,26127. .26330,
26585. .26783,26941. .27842,27936. .27991,28208. .28413))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translation="MDSVDRKINNTTQSSPSPKPSKFSVYRNPAALAASTANIRP
SKSVLFIPLLSASAPLISFWAGEKWLNTALTPGKISQEAAYVTKVAQGLVTLEFC
IGAMWLSKGISLRHAGETKSPSKETKQDFSLNSRQELLGILKKADQSVLFC
PKRPAKPAQSLPELPIIHQTLVSAHSGSGGDKLNSRNGSQISFSTPSKQWMS
PSMYLPSSSPVSNRLSSGQKAPSPWSSRRSADKITTTEOLEQLAEIDIKETIE
SAGKMRTPPTKPSFAMASPTVSGTSGAGSTRSTPLRPVMSPGAKFTTPPKGE
GDPNPNLSLRATEGFIHLGVYQIEDWRDLRQWSSVLLKPLLNKVTSHIQVMT
ASKLGNVTVSQVSDIPTNGTATTALPVDRTKQWQPSYSLDEDLHLQURANLVQAI
DASHQKLRQEQOQQOQQOQOQALIPVQECVDALSHRRLLQGLMKGEWVGLLP
RSSIPADYTVQIRALAEQCVKNEYNGADTRKNKWSLEPPTDHLHLLFCAFP
LEHPKMWLDPSSYTGTOASKNPLFLGVLPKPERPPEKVIYVGVSPVTLHPGACVL
AVDKQSPPTALYWDKVKQFTLQRTALWDSMLLIChRIKVGVGVRGNLGSALN
ILQVDSDDTD"
complement (join 29418. .29491,29579. .29620,29858. .29942,
30386. .30805))

/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translation="MVSSKPKPKKARSDVVVNKASGRSKRSSGRSTKTKSNKVNIVK
KPEIYEISSESSSDSVEEAIKRGDEAKSGVYKRGNGKSVGIPPTKSKNREDDGG
ABDAKIKFPMNRIIRIMSDNSAPQIMQDAFLVFNKATEMFIERFSEAYDSSVYDKK
KFTHYKHLSSVNSDQRYEFLADSVPEKAKAAALBEWERGMTDAG"
Alignment Scores:
Pred. No.: 1,88e-229 Length: 120965
Score: 2971.00 Matches: 616
Percent Similarity: 68.40% Conservative: 1
Best Local Similarity: 68.29% Mismatches: 4
Query Match: 92.38% Indels: 284
DB: 8 Gaps: 3
US-10-018-929c-3_COPY_650_1270 (1-621) x AC026875 (1-120965)
Qy 1 ArgAlaAspAlaPheLeuPheGlySerSerLeuAenProSerHisAspValIleHis 20
Db 63512 CGAGCTGATCGGTTTATTTCTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAGCAC 63453
Qy 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
Db 63452 GTAGAGAAGATAAAATCGAGTCATGTTCTTGAAGAAGAACTAAGATATTTCCGATTGTACTCA 63393
Qy 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAenMetArgGlnAenLys 60
Db 63392 GTATGTACAGTTGAAGAAAAGCCCTGATTTCTGCTAGGCAAAATAAGCGCAAAATAAA 63333
Qy 61 AlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyr 80
Db 63332 GCTGTAGAGAACCTAAACCCCTCTCTCACGCACGCACTGCTCATGTGGGGGCGTCATAC 63273
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 63272 TTATTTGATAAATCGGATCATATTTTACAGCAGTGAAACTCCAGATTCCAGGAGTTTCAATTT 63213
Qy 101 GluLysSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 63212 GAACAATCTATATGAGCGGGCGGATTCATGAATCTTCGTCATACATCTTCTTCCAAAGGT 63153
Qy 121 GlyGluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 63152 GGAGAGAANAATGAGTCAAGCTGTGCTACTTTTGGAGGCCAAGCATGCTCAGGGAAT 63093
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 63092 TACAGCAGTGAATCTACTCTTATTTGGTGAAGACCATATTAAGTTGTGATGAAGAGAGT 63033
Qy 161 ProAsnIlePheThrPheSerLysLeuLeuGlyGlyLysAenProMetTyrLysTyrProSer 180
Db 63032 CCAATATATATTTGGTCAAAAGCTGTGGGGGGGAAAAATCTCTATGTGGAATAATCCCTTCA 62973
Qy 181 AspThrProGlnArgAenArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 62972 GATACTCCCCAAGGAATCCAAAACGATTCAGTTATTTTGGGGGTTCTTGAAGCAGATCCC 62913
Qy 201 LysThrGlyAspGlyGlyAenAlaLysLysArgLysLysAlaSerAspAspValThrAsp 220
Db 62912 AAAATGCGCATGGTGAAATGCAAAGAGCAAGAAAGGCTTCTGTGATGTGCTCACTGAT 62853
Qy 221 ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 62852 CCCCCGGTCACTGATCCGCCAGTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 62793
Qy 241 Met----- 241
Db 62792 ATGGGTHAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTGTTCAATTGATTATCAACTA 62733

QY 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
 Db 62732 CTTTGCTGACTATCTTTCCTACAGGGCT-TTGGAGTCACAAAAGTCATAA CACTCCAG 62674
 QY 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu 273
 Db 62673 TCATCATGTAAATCTCTGCTGACAGATGGTACATTTGGATGGAAATGATGCTTTTGGCTTG 62614
 QY 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTTP 293
 Db 62613 TATTTCTATGGGCGCCCATATCTCTGGAAATCCAGAGGATATGTAGCTAGTCAAGATTGG 62554
 QY 294 GlyLysIleProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetAla 313
 Db 62553 GGGAAAATACCGATGNAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAAGATGGCA 62494
 QY 314 LysLeuCysGlnValLeuHisLeuSer----- 322
 Db 62493 AAACCTTGCCCAAGTTTGGCATCTTTC-AGTAAAGTGGGCTTTTTCACCTCCACAACCTTATT 62435
 QY 322 ----- 322
 Db 62434 TTAGCCTTGCAATATGCTTATATATAGCTGANTGCAACTGTAGTTTACCTGATTTCCCTG 62375
 QY 322 ----- 322
 Db 62374 TTACAGCCAAATGTGAGAGTTTATCTTCAACTATATCATCTCGTTTAAAGCATATTTTA 62315
 QY 322 ----- 322
 Db 62314 TTCTTTATATCTGGCTTCGTTTACCAGTCACTGTAAATGAGCAACTGCTGCACAAAAC 62255
 QY 322 ----- 322
 Db 62254 AGTAGTAGTTATGTGCTCATGTCAATTCATTTGTTTATTGAAGCAAGAAATTTCTGTCT 62195
 QY 322 ----- 322
 Db 62194 ACTTTACATGATCCATCTGTGGGAGTATATACTATATATAACCTTAGGCTTTGTACCT 62135
 QY 322 ----- 322
 Db 62134 GGCTGATCAAAGACATGTCAAAGTTTATCTGTTGCTGCTGTTGGTATAGAACTAATACAG 62075
 QY 322 ----- 322
 Db 62074 TGTCTGATGCTATTTTAAAGTAGTCTTTATGCTTTACATATATGGCTAATAGATGTTTCCG 62015
 QY 322 ----- 322
 Db 62014 CTGTCGTGTCATATACCTTCTGTGATTTATCACGGTGTCTCGTCTATCAAAATTTGTACTAA 61955
 QY 323 -----AspAlaCysThr 326
 Db 61954 AAGGTATTTTGAATGTGATGTTGTTTAAACAGATTAATTTGTTTTCAGGATGCTTGCA 61895
 QY 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
 Db 61894 AGCATGGTTCGGAAATTTTCTCGAATATGTTTATTGAAAATCACCGAATCTACGAAGAGCCA 61835
 QY 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
 Db 61834 GCCACTACTTTTCAGGATTTCCAGATAGCCCT-GGTATGACAGCATTTACTTTTGATTAAT 61776
 QY 357 ----- 357
 Db 61775 TATGCAATGTTTCTTCATCATCTGCTTGTGTTTGAAGATGCTTCAGAGGCACTCC 61716
 QY 357 ----- 357
 Db 61715 TTTAGTTTAACTTTCCAATCATAGGATTCAAATATCCATTAATGCGCTTTGATCGCTG 61656
 QY 358 -----SerTTP 359

Db 61655 CATAATATATGATAGTTGACATACCTGAATACGTTGTTAATAATGCAATTTTCAGAGTTGG 61596
 QY 360 IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
 Db 61595 ATTGCAAGCTTGTGTGTAAGCAAAATCTTAGCCACAAAAGAAATCTCTGGTCCGTGCAAAAT 61536
 QY 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
 Db 61535 TCTGAATTTAGCTTTCAATATGCTCTAGAGTAGAGTGAGTTATATTATTTCGATATTGTCC 61476
 QY 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
 Db 61475 TGCATGAAGAGTCTGTTCTCTGGAGCATACACAAGGTTTGCAGTTTCGATTTGTTGTTACT 61416
 QY 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr 439
 Db 61415 AATTTCTAAAACAGTCAGTGGTTAGCACAAAACCTAGTAAATGAAAGTCTCTCAGGGGTACA 61356
 QY 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
 Db 61355 GTCCGTGACGAAAAGATTTAATACGAAGTCGATGCGAAATAGCTCAGAGGATGAAAGTGC 61296
 QY 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
 Db 61295 ATGACTGAGAAGAGATGTAGCCATTATAGCACAGCAACAAGAGATATCGAAAAGACTATT 61236
 QY 480 SerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
 Db 61235 AGTGGCATAAAAAAGAAATACAGAAGCAAGTGTCAAAAGCTTGTACAAGAGCATGAGAA 61176
 QY 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
 Db 61175 AAGAAAATGAGCTGTTAATATATGATGACACAGACAGCAAGAACTTGAACCTAGTAAA 61116
 QY 520 SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
 Db 61115 AGTGTGGAGCAGCAGTAATTCGTATTACTCTGTTTCAGGACCACTACTCAAGTGGGTGAT 61056
 QY 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn 559
 Db 61055 CTCAAACTGCTGGATCATATAATTAAGAAAGAAAGTTTGTAGAAATCAAAAGTGAAGAAAAT 60996
 QY 560 GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
 Db 60995 GAATGCTCAAAGCTCTGGAGCAAAATGCAGATGTTGCCAAGAGAAAGTTGGCTGAGGAT 60936
 QY 580 GluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLysLysValCysValPro 599
 Db 60935 GAAGCCTGTTGGATTAAATCGGATAAAGAGCTGGGCGAGCTAAAATTAAGATTTGTGTTCCC 60876
 QY 600 IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro 619
 Db 60875 ATTCAAAGTGGCAATAACAAGCAATTTTAGTGGTTTCATCAAAACATTTCCCAAAATGCTCCT 60816
 QY 620 AspVal 621
 Db 60815 GATGTA 60810
 RESULT 6
 ATH552638
 LOCUS
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone
 340E12.
 ACCESSION AJ552638
 VERSION AJ552638.1 GI:29368785
 KEYWORDS left border; T-DNA flanking sequence.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1

AUTHORS Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M., and Lecharny, A.

TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites

JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)

MEDLINE 22363535

PUBMED 12446565

REFERENCE 2 (bases 1 to 695)

AUTHORS Balzergue, S.

TITLE Direct Submission

JOURNAL Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment (s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

FEATURES Location/Qualifiers

source 1..695

organism "Arabidopsis thaliana"

mol_type "genomic DNA"

cultivar "Wassilewskija"

db_xref "taxon:3702"

clones "340E12"

clone_lib "Arabidopsis thaliana T-DNA insertion lines"

misc_feature 1..695

note "T-DNA flanking sequence left border"

ORIGIN

Alignment Scores:

Pred. No.: 6.77e-75 Length: 695

Score: 1033.00 Matches: 203

Percent Similarity: 87.88% Conservative: 0

Best Local Similarity: 97.88% Mismatches: 1

Query Match: 32.12% Indels: 28

DB: 8 Gaps: 1

US-10-018-929C-3_COPY_650_1270 (1-621) x ATH552638 (1-695)

Qy 104 ILeMetaspGlyValIleHisGluPheSerSerIleLeuSerSerIleGlyGluGlu 123

Db 2 ATATTGACGCGCGTATTCATGAATTCGTCCTCATCTTTCTCCAAAGGTGGAGAGAA 61

Qy 124 AsnGluValIleGlyLeuGlyLeuGluAlaGlyHisAlaGlnGlyThrTySerSer 143

Db 62 AATGAGTCAGCTGTGTCTACTTTTGGAGCCCAAGCATGCTCAGGGAATTCAGCAGT 121

Qy 144 AspSerThrLeuPheGlyGluAspHisIleLeuSerSerAspGluGluSerProAsnIle 163

Db 122 GATTCTACTTATTGGTGAAGACCATATTAAAGTTGTCAGATGAAGAGATCCAAATATA 181

Qy 164 PheTrpSerIleLeuGlyGlyValIleGlyAsnProMetTrpIleTyProSerAspThrPro 183

Db 182 TTTTGGTCAAAAGCTGTTGGGGGGAATAAATCTATGTGGAAATACCCCTTCAGATACTCCC 241

Qy 184 GlnArgAsnArgIleGlyValGlnTyPheGluGlySerGluAlaSerProIleGlyThrGly 203

Db 242 CAAAGGAATCGAAACAGATTCAGTATTTTGGGGTCTTGAAGCGAGTCCCAAAATCGGC 301

Qy 204 AspGlyGlyAsnAlaIleGlyValIleGlyValIleGlyValIleGlyValIleGlyVal 223

Db 302 GATGGTGAATGCAAGAACGAAAGAGAGCTTCTGATGATGTCACATGATCCCGGGTC 361

Qy 224 ThrAspProValAspAspGluArgGlyAlaSerGlyIleGlyAspHisMet----- 241

Db 362 ACTGATCGCCAGTAGATGATGATGAAGAAAGAGCCCTCTGGGAAGGATCATCATGGTTAAA 421

Qy 241 ----- 241

Db 422 ATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTACTTTGCTGA 481

Qy 242 -----GlyAlaLeuGluSerProIleValIleThrLeuGlnSerSerCys 256

Db 482 CTATCTTTTCCPACAGGGGCT-TTGGAGTCACCAAAAGTCATAACACCTCCAGTCATCATGT 540

Qy 257 LysSerSerGlyThrAspGlyThrLeuAspGlyValAspAlaPheGlyLeuTySerMet 276

Db 541 AAATCTTCTGTCACAGATGTACATTCGATGGAATGATGCTTTTGGCTGTATTCATG 600

Qy 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyValIle 296

Db 601 GGCAGCCATATCTCTGGAATCCAGAGAGATATGTTAGCTAGTCAAGATTGGGGGAAATA 660

Qy 297 ProAspGluSerGlnArgArgLeuHisThrVal 307

Db 661 CCGGATGAATCACAGAGAGGAGCTCCACACTGTT 693

RESULT 7

AK068790

LOCUS AK068790 4027 bp mRNA linear PLN 24-JUL-2003

DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013163J20, full insert sequence.

ACCESSION AK068790

VERSION AK068790.1 GI:32978815

KEYWORDS FLI CDNA; CAP trapper.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE 1

AUTHORS The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team; Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Otonoda, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, R., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Oato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayaishizaki, Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 4027)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayaishida, K., Hayaishizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohta, M., Ohtsuki, K., Oka, H., Ooka, H., Ootani, N., Ota, Y., Otonoda, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,

Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
 Yoshimura, A.
 Direct Submission
 Submitted (05-DEC-2001) Shoehi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp.
 Tel: 81-29-838-7007, Fax: 81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica
 rice.
 URL: <http://cdna01.dna.affrc.go.jp/cDNA/>
 NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
 Yamamoto, M.
 FAIS Genome Sequencing & Analysis Group: Otono, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Niihara, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
 Ota, Y., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1. 4027
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="J013163J20"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3 01e-32 Length: 4027
 Score: 521.50 Matches: 132
 Percent Similarity: 50.12% Conservative: 80
 Best Local Similarity: 31.21% Mismatches: 137
 Query Match: 16.22% Indels: 75
 DB: 8 Gaps: 10
 US-10-018-929C-3_COPY_650_1270 (1-621) x AK068790 (1-4027)
 QY 3 AspAlaPheLeuLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
 Db 2899 GATGCCATCATATATATTGATGATGGATCCCAACCAATGACTTGAGAGTCTCCAG 2958
 QY 23 LysLeileLysleGluSerCysSerGluArgThrLyslePheArgLeuTyrSerValCys 42
 Db 2959 AGGCATGATAGAGTCTCAATCTGAAATGTGTGCTATTTTCGCTTGTATTCGTCGTG 3018
 QY 43 ThrValGluGluLysAlaLeuLeuLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
 Db 3019 ACGGTGGAGAAAGAGCTCTTATCTAGCGAAGCATGATCATATCTTGGACCAACGTC 3078
 QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82

3079 CAGAATGTAATGCCTATCGTGAGCCATTCCTCGCTTAGTTGGGTCGATCGTTCTCTTC 3138
 QY 83 AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
 Db 3139 ATATAGACTTGGAGAGTTCCAAAAGCATGACTACTCCAGTAAGAAATCTCTAGGATGATGGT 3198
 QY 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
 Db 3199 CTC-----TTCATGAATATGATTTTGGAGTTTGCAGCGAAACTATCC 3243
 QY 118 SerLysGlyGlyGluGluAenGluValLysLeuCysLeuLeuGluAlaLysHisAla 137
 Db 3244 ACCAAGCTTGAAGCCAGCACTAAATGAAAATGCAATCATATCTCGAGCTCAGCAGAGT 3303
 QY 138 GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSer 156
 Db 3304 GGATCGTTTATTTCTAGATATGCTGTTTAAAGGAGGAGGAGGAATATCTCGAGTT 3363
 QY 157 AspGluGluSerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
 Db 3364 GATGGTGAATTCGCAAGTTCTGGACCTTCTGGTCAAAATTACTAGTGGGAGTCTCCT 3423
 QY 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
 Db 3424 CATGGCAATATATATCTGAGCCAGTGCAAAAGGACCGTAGGAAGATACAAAATATATGAA 3483
 QY 195 GlySerGluAlaSerProLysThrGlyAspGlyLysAsnAlaLysLysArgLysAla 214
 Db 3484 GACCAATGAGAAATCTCGCAGAGAAACTGATGAAGCAATATGAACATGAGAAAATC 3543
 QY 215 SerAspAspValThrAspProArgValThrAspProValAspAspGluArgLys 234
 Db 3544 GGAGAA-----ATCATGGAT-----ATGATGCTGTG----- 3558
 QY 235 AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValleThrLeuGlnSer 254
 Db 3559 -----TCATCTCCGAAGATCCTACCTGTGCAAGAC 3588
 QY 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
 Db 3589 -----ATGATGCTGTG----- 3600
 QY 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAsp----- 292
 Db 3601 -----CTGCCTGAAATAGTACAGCATCCAGTCTCATGAA 3636
 QY 293 -----TrpGlyLysIleProAspGluSer-----GlnArg 302
 Db 3637 ACATCAGTTGATGACACACTTGGCAAGAGCTAGGGGCGAAGAGCCTTCAGGGCACACAGAA 3696
 QY 303 ArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSer 322
 Db 3697 GGTCTTCACACCACTCAAGCCAGAGCTGTCAAGATTGATGAATTGTGGAAATGCCG 3756
 QY 323 AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIle 342
 Db 3757 GAGAGTGTCAAATGTTTATGTGAGGAATTACTTGATTACATTTTGAAGATCATCAAGTC 3816
 QY 343 TyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAla 362
 Db 3817 AGTCAGGAGCCAAAGGCGATATTGCGATGCATTCACATTTGCCCTGTGTGCGGTGCTG-- 3874
 QY 363 LeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeu 382
 Db 3875 CTTCTCTCAAAGCATTAAGATCAATCGAAGAGAGTCACTTCCCTCTTCTGTGAGAAACTTG 3934
 QY 383 AlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLys 402
 Db 3935 AACTAGTGGTCCGATCAAGTGTCTCGCTGAATATGTTTATGAGAAGCTTAAGGATCCTCAAG 3994
 QY 403 SerLeuPhe 405
 |||

Db 3995 AAAAGTTT 4003

RESULT 8
AX063609
LOCUS
DEFINITION Sequence 28 from Patent WO0100801.
ACCESSION AX063609
VERSION AX063609.1 GI:12541333
KEYWORDS
SOURCE
ORGANISM Brassica oleracea

REFERENCE 1
AUTHORS Habu, Y., mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 28 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

FEATURES
source 1..467
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="unassigned DNA"
/db_xref="taxon:3712"
/note="seq1-43"

ORIGIN
Alignment Scores:
Pred. No.: 2,47e-31 Length: 467
Score: 494.50 Matches: 97
Percent Similarity: 84.09% Conservative: 14
Best Local Similarity: 73.48% Mismatches: 12
Query Match: 15.38% Indels: 9
DB: 6 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x AX063609 (1-467)

Qy 111 GluPheSerSerLeuSerSerLysGlyGluGluAenGluValLysLeuCysLeu 130
Db 1 GAATTCCTCGTCCATACCTTCTCCGATGTTGGAGAGAAATGAAGCAAGCTGTGCTA 60
Qy 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTyrsSerSerAspSerThrLeuPheGlyGlu 150
Db 61 CTTTGTGAAGCCAGCATGCTCAGGGAAGTTACAGCCTGATGCTACTCTATTGTTGTA 120
Qy 151 AspHisLeuLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuGly 170
Db 121 GAACATGTCAAGTTATCAGAT---GAAAGTCCAATATGTTTGGTCAAGCTGTGAGT 177
Qy 171 GlyLysAsnProMetTrpLysTyrsProSerAspThrProGlnArgAsnArgLysArgVal 190
Db 178 GGAAGAACCTATGTGGAAATACTGTTCGATACTCTCAAGAGAGTCGAAAAAGAGTA 237
Qy 191 GlnTyrsPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLys 210
Db 238 CGGCATCTTCAGGGCTATGAGGAGACTACCAAGTTGGCAATGGCGGAAACTTAAGAAG 297
Qy 211 ArgLysLysAlaSerAspAspValThrAspProArgValThrAspProValAspAsp 230
Db 298 AAAAAGAGGCTTCAGATGATGTCA-----GTAGATAAC 333
Qy 231 AspGluArgLysAlaSerGlyLysAspHisMetGly 242
Db 334 GCTGAGAGAAAGCTCTGGAAAGGATCAGATGGGT 369

RESULT 9
AX063611
LOCUS
DEFINITION Sequence 30 from Patent WO0100801.
ACCESSION AX063611
VERSION AX063611.1 GI:12541335
KEYWORDS

SOURCE ORGANISM Brassica oleracea

REFERENCE 1
AUTHORS Habu, Y., mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 30 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

FEATURES
source 1..515
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="unassigned DNA"
/db_xref="taxon:3712"
/note="2-33"

ORIGIN
Alignment Scores:
Pred. No.: 1,91e-29 Length: 515
Score: 472.00 Matches: 100
Percent Similarity: 60.87% Conservative: 26
Best Local Similarity: 48.31% Mismatches: 39
Query Match: 14.68% Indels: 42
DB: 6 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x AX063611 (1-515)

Qy 387 SerArgValGluValAspTyrsIleTyrsSerIleLeuSerCysMetLysSerLeuPheLeu 406
Db 1 TCTAGAGAGAGGGTGGATTATATATCTTTCTGACTGCTGAGAGAGCTATTCTGTG 60
Qy 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
Db 61 GGGCGCACACAGGTTTCCAA----- 81
Qy 427 SerThrLysLeuValAenGluSerLeuSerGlyAlaThrValArgAspGluLysIleAen 446
Db 81 ----- 81
Qy 447 ThrLysSerMetArgAenSerSerGluAspGluGluCysMetThrGluLysArgCysSer 466
Db 82 -----GAAAGGGTGAAGATGCTGCTGAGAAAGAGGTAGC 120
Qy 467 HisTyrsSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTyr 486
Db 121 CATATAGCTCAGTAACCAAGGATGTTGAAAGACTATTAGCGACATCAAAAGAAATGC 180
Qy 487 LysLysGlnValGlnLysLeuValGlnGluHisGluLysLysMetGluLeuLeuAen 506
Db 181 AGTAAGAGCGCTGCATAGCTTTGTACAAACCTCGAGGAAGAAAAGATGGACCTGATGA 240
Qy 507 MetTyrsAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIle 526
Db 241 AGGAATCTCTCAAGAAAGCAGGAACCTTCAGAAATGTAAAGGGTGAAGCATCATTTAT 300
Qy 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
Db 301 CGTGTCACTATTTCAGGTATAAATACTCAGAGCTTACATGCTCTCCAACGGCTGGA 360
Qy 545 HisAenTyrsGluArgLysPheAspGluIleLysSerGluLysAenGluCysLeuLysSer 564
Db 361 TGTACTTTTGAAGAAAGTTTGATGATCTCAAGAGGAGTTGGATGAATGCTTGAAGAT 420
Qy 565 LeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
Db 421 TTAGAGCAATAAACCAGGCTGGAAGAGAAAGTTGGCTGAAGATGAAGCTGTTGGATT 480
Qy 585 AsnArgIleLysSerTrpAla 591
Db 481 AGTCGGATAGAGAAATGGCA 501


```

RESULT 10
AX063607/c
LOCUS AX063607 519 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 26 from Patent WO0100801.
ACCESSION AX063607
VERSION AX063607.1 GI:12541331
KEYWORDS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 Habu.Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
AUTHORS Gene involved in epigenetic gene silencing
TITLE Patent: WO 0100801-A 26 04-JAN-2001;
JOURNAL Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source
1..519
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="unassigned DNA"
/db_xref="taxon:3712"
/notes="seq1-23"
ORIGIN
Alignment Scores:
Pred. No.: 5,31e-27 Length: 519
Score: 442.00 Matches: 89
Percent Similarity: 84.82% Conservative: 6
Best Local Similarity: 79.46% Mismatches: 13
Query Match: 13.74% Indels: 4
DB: 6 Gaps: 1

US-10-018-929C-3_COPY_650_1270 (1-621) x AX063607 (1-519)
QY 1 AqAlaaspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
Db 324 CGCGCTGAGTCTTTATCTTTTGGAGCAGCTGAATCCATCGCATGACGTTAAGCTC 265
QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
Db 264 TTAGAGAGAGATAAAGGTTGAGTCTGTCTGAAAGAACTAAGATTTTCGGTTGTACTCA 205
QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
Db 204 GTGTTTACAGTTGAAGAAAAACCCCTGATCTTGCT-----AGGCAAAATAAG 157
QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyr 80
Db 156 CCTCTAGAGAACCTTAACCCCTCTCTCACACATGCAATGCTCATGTGGGGGGCATCATAT 97
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 96 TTATTGTATAAGCTGGAACACTCTCCATGGGAGTGGAAACCCAGGTTCCAGGATTCATTT 37
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
Db 36 GAACAGTCTTATAAGGATGCCGTAANNACGGAATTC 1

RESULT 11
AP005621/c
LOCUS AP005621 184118 bp DNA linear PLN 27-FEB-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
PAC clone: P0681G01, complete sequence.
ACCESSION AP005621
VERSION AP005621.2 GI:28570002
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Sasaki,T., Matsuoto,T. and Katayose,Y.
AUTHORS Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
clone:P0681G01
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 184118)
AUTHORS Sasaki,T., Matsuoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
[E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468]
COMMENT On Feb 26, 2003 this sequence version replaced gi:22091093.
The orientation of the sequence is from 77 to 586 of the PAC clone.
FEATURES
source
1..184118
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="6"
/clone="P0681G01"
ORIGIN
Alignment Scores:
Pred. No.: 7,55e-20 Length: 184118
Score: 397.00 Matches: 178
Percent Similarity: 32.47% Conservative: 134
Best Local Similarity: 18.52% Mismatches: 255
Query Match: 12.34% Indels: 395
DB: 8 Gaps: 22

US-10-018-929C-3_COPY_650_1270 (1-621) x AP005621 (1-184118)
QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 19304 GATGCCATCATCATATATTGTAGTATTGGAATCCAAACCAATGACATTGAGAGTTCTCCAG 19245
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 19244 AGCATGATATAGAGTCTCAATCTGAATGTGTGCTTATTTTCGTTGTATTCGTTGT 19185
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 19184 ACCTGGAGAAAGAACTCTTATCTAGCAAGCATGATCATATTCTTGACACCAAGCTC 19125
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyrLeuPhe 82
Db 19124 CAGAAATGATGCTATCGTGAGCCATTCCCTGCTTAGTTGGGTGCATCGTTCTCTTC 19065
QY 83 AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
Db 19064 AATAGACTTGAGAGATTCCAAAAGCATGACTACTCCAGATAAAGATTCTGAGGATGATGTT 19005
QY 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
Db 19004 CTC-----TTCATGAATAATGATTTTGTGAGTTTGCAGGAAACTATCC 18960
QY 118 SerLysGlyGlyGluGlnAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAla 137
Db 18959 ACCAAGCTTGAAGCCAGCACTAAATGGAATGTCAGTCATATCTCGAGCTCAGCAGAGT 18900
QY 138 GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSer 156
Db 18899 GCATCGTTTTTATTCTAGAGATATTGCTTTATATAGCGAGAGGAGGAGATATCTCGCAGTT 18840
QY 157 AspGluGluSerProAsnIle-----PheTyrSerLysLeuLeuGlyGlyLysAsnPro 174
Db 18839 GATGGTATTTGCCAAAGTTCTGGACCTTCTGTGTCAAATTTACTAGTGGAGGCTCTCCT 18780
QY 175 MetTyrLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194

```


Db 18779 CATGGCAATATATCTGAGCCAGTGCAGAGAACCTAGAGAGATACAAAATATGGAA 18720
Qy 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLeuArgLysLysAla 214
Db 18719 GACCAATGAGAAATACCTGCAGAGAAACTGATGAAGCAATAATGAACGGTAGAAAATC 18660
Qy 215 SerAspAspVal---ThrAspProArgValThrAspProProValAspAspGluArg 233
Db 18659 GGAGAAATCATGGAAATCATCTCCGAAGATCCTA-----CGTGCAAGACAATGAT--- 18609
Qy 234 LysAlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValLleThrLeuGln 253
Db 18608 -----GCTGTGCTGCCTCGAA 18594
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGly----- 267
Db 18593 AATAGTACAGCATCCAGTCTCATGAACATCAGTTGATGATGACACTTGGCAAGAGCTAGGT 18534
Qy 268 -----AsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle 280
Db 18533 ATCCAACTGTAACAGTAATTAACAGCTCTTATGCTTTATGTGATATGCTACATTTTTT 18474
Qy 281 SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer 300
Db 18473 TATAATACACTCCATTTTGTAGGGCGAGAAAGCCTTCAGGCG-----ACA 18429
Qy 301 GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis 320
Db 18428 CAGAAAGTCTTCACACCCTCAAGCCAGAGCTGTCAAGTTGTATGAATTTGTGAA 18369
Qy 321 LeuSer----- 322
Db 18368 TTGCGGCTACTGCCAAGTATATAAATAATGCTTGTCTTTATGTTATTAATTTGAGAAA 18309
Qy 322 ----- 322
Db 18308 AATATTTAGTTCTACTGCAAGTACAGGACATAATCTGCTCTTTTAGCATTTCAATTAATA 18249
Qy 322 ----- 322
Db 18248 GCTAATATTTTGTAGTTTCAACAGCAGAGTCCACACCTTGTGAAGAGTTATTTATCAA 18189
Qy 322 ----- 322
Db 18188 GTTGGATCTTGATAACTTTTCTCTTCATACGGAACAATTTCTTGACTCTTAGTGTG 18129
Qy 322 ----- 322
Db 18128 TCATTTTGATCCATATATGTTACTTCTCCTCGTTTCACAATGTAAGACTTTCTAGCATTCG 18069
Qy 322 ----- 322
Db 18068 CCACATTCATTTAGTGTATATGAATCCAGACATATGATGTGTGTTTAGATTCATTAACAT 18009
Qy 322 ----- 322
Db 18008 ATATATGATGTGGCACTACTAGAAAGTCTTACATTTGTGAATGGAGGAGTATTAGTT 17949
Qy 322 ----- 322
Db 17948 GCTACTATATCCTGCTTTTAGCTCAGGTTATGAGCTTTGATATCACCTACTCCTCCGCC 17889
Qy 323 -----AspAlaCysThrSer----- 327
Db 17888 CCATAGAAAACCACTAGTACTGATGCTGTCTAGATTCGTAGTACTAGGATGTGTCA 17829
Qy 327 ----- 327
Db 17828 CATCCGCTACTAGGTGTGTTTTCTATGGACGAGGAGTGCATGTTTCTACATGCTTCC 17769
Qy 327 ----- 327
Db 17768 AATATGAATTTATGACGACAGCTGTGTGCTGCGACTCTGCTGTAAAGTGTGCTCTCT 17709

Qy 327 ----- 327
Db 17708 CTTCTGTAGAAATTTATGCGAGCATGCTGTACCTAAGTTTCAGTTTAAGCGAACTTTTCC 17649
Qy 327 ----- 327
Db 17648 TGAAGTCACTTGTATCTCTAGTTAAATAATAAAGCTGATTTGGATGCTTGGATGCTTGTAT 17589
Qy 328 -----MetValGlyAsnPheLeuGluTyrValIle 337
Db 17588 AAAACCTTTTTCAGGAGACTCTCAATGTTTATGTGAGGAATTAATCTGAT-TACATTTTG 17530
Qy 338 GluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeu 357
Db 17529 AAGAATCATCAAGTCAGTCAGAGGCCAAGGCCATATTGCATGTCATTCAACATTTGCCCTG 17470
Qy 357 ----- 357
Db 17469 GTTAGATTATAAAAAATATTTTATAGTTTTTATAGTTTTTGTCAATTTGTCATGCTAATGCTATT 17410
Qy 358 -----SerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSe 370
Db 17409 ATTTTTTCCATGCTCTTTTCAGTGTGCGTCTGCTCTCTCTTAAAGCATATAGATCAA 17350
Qy 370 rHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValG1 390
Db 17349 TCGAAGAGAGTCACTTGCCCTTTCTGTGAGAAACTTGAACATATGAGTCGATGAAGTCT 17290
Qy 390 uValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThr-- 409
Db 17289 CGCTGAATATGTTTATGAGAAGCTTAAGGATCCTCAAGAAAAAGTTTTCGCTAGACGAAG 17230
Qy 410 -----GlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerVa 425
Db 17229 TGAACAACAGCAGCAGCCACTCTACTCCAGTAAATAACACATCACTTACAAACAACA 17170
Qy 425 lValSerThrLysLeuValAsnGluSer-----LeuSerGlyAlaThrValArgAspG1 443
Db 17169 GACCTCACCGAAATTTGAGAAGTGACGGATCAATTTGTTCATCAGGTAAACACATTTGATGG 17110
Qy 443 uLysIleAsnThrLysSerMetArgAsnSerSerGluAsp-----GluGluCysMe 460
Db 17109 TGATTTGGAAAAATGTTTCATCATGAAGAGGCTCCACATGATATCTTGACCGAGGAGATGAT 17050
Qy 460 tThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSe 480
Db 17049 ATTAGAACAGAG-----GARTTGATATC 17026
Qy 480 rGlyIleLysLysLysTyrLysLysGlnVal-----GlnLysLeuValGlnGluHisG1 498
Db 17025 TGTTCAGAAACTCATAGAGAAGAACATGTTTTTAAGGGATGAACCTTCTTGAAGAATCAC 16966
Qy 498 uGluLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLys----- 514
Db 16965 AGAAAAAGAAATTAATTTAATTAACATGCTTTCTCCTTAAGAGAAAAGATATTTACGGA 16906
Qy 515 -----LeuGluThrSerLysSerValGluAlaAlaValI1 526
Db 16905 TAAACAAGGAAACGAGACCACACTGTTAGACATCCACAGCAAGAAAGGATGGCAAGCT 16846
Qy 526 eArgIleTyrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHis 546
Db 16845 GCAAGAACATGCAAT-----TTGTCGTGGAACTCT 16813
Qy 546 nTyrGluArgLysPheAspGluLysSerGluLysAsnGluCysLeuLysSerLeuG1 566
Db 16812 TCGTAAAGGTCACATTCAT-----TCAGAGGACGGGATGCCACGGTGAAGATAAT 16759
Qy 566 uGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnAr 586
Db 16758 TGAATGTTTACTTTGCTTCTGTATGCAATTTTGAATCACATGATGATGCCAGCATAACA 16699

```

QY      586 gilelys-----SerTpAlaAlaLysLeuLysValCys---ValProil 600
      ::::|||||
Db      16698 GTTGAGATCAGCAATCACTTCATGGAACAGAGTTACAAATGAGGAATTTTCCT 16639
      ::::|||||
QY      600 eGlnSerGlyAsnAsnLysHisPheSerGlySerAsnIleSerGlnAsnAlaProAs 620
      :|||::|
Db      16638 TCAGCAAGCAAAATCTGGCCACTTAGATCGTAGCTTTTGATCAACAAATTCCTTTACCAGA 16579
      :|||::|
QY      620 p 620
      :|||
Db      16578 T 16578

RESULT 12
AP005005
LOCUS   AP005005              150597 bp      DNA      linear      HTG 28-MAR-2002
DEFINITION   Oryza sativa (japonica cultivar-group) chromosome 2 clone P0513D05,
*** SEQUENCING IN PROGRESS ***.
ACCESSION   AP005005
VERSION     AP005005.1 GI:19773545
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1
AUTHORS     Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE       Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
            clone:P0513D05
JOURNAL     Published Only in Database (2002)
REFERENCE   2
AUTHORS     Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE       Direct Submission
JOURNAL     Submitted (27-MAR-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source           1..150597
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="2"
                     /clone="P0513D05"

ORIGIN
Alignment Scores:
Pred. No.:          1.01e-14          Length:          150597
Score:              332.50            Matches:         199
Percent Similarity: 30.64%            Conservative:    108
Best Local Similarity: 19.86%          Mismatches:      269
Query Match:        10.34%            Indels:         428
DB:                  2                Gaps:           28

US-10-018-929C-3_COPY_650_1270 (1-621) x AP005005 (1-150597)

QY      3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
      :|||||
Db      137548 GATGCCATCATATATATGTTCTGTGACAAACACCCATTAACGACTTGAAGGCTCTTCAA 137607
      :|||||
QY      23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42

```

```

Db      137608 AACATCAAGATAGAGTCACAAATTTGAGCGTGCAGCATTTTTCGTTGTACACTCCTTTT 137667
      :|||||
QY      43 ThrValGluGluLysAlaLeuIleLeuAlaAArgGlnAsnMetArgGlnAsnLysAlaVal 62
      :|||||
Db      137668 ACAGTTGAGGAGAAGAGTCTTGTCTTGGAGGAGCGGCAATGTTATTATTATAACAACATA 137727
      :|||||
QY      63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
      :|||||
Db      137728 CAAGATCTAAGGACTAGTTTAAACATTCCTTAATCAGATGGGCTGCTGCATTTCTTCTTC 137787
      :|||||
QY      83 AsPlysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhePheGluGln 102
      :|||||
Db      137788 AGTAGACTTGATCAGGTCCAGCAAGATGACCATCAAGTAAAAAGTTCTCGAAATGGAGAGA 137847
      :|||||
QY      103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlyGlu 122
      :|||||
Db      137848 CACTTTATTGATGAAGTAATCGTTGAGTTCTTAAACAAAGTTATCCCACTGTTGAAGAC 137907
      :|||||
QY      123 GluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
      :|||||
Db      137908 AGCATTGAAGTACACAGGAAGTCCATAAGCAAGCTAATATGATGGGGAACTATATTCA 137967
      :|||||
QY      143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
      :|||||
Db      137968 AGAAATATTACTCTGATGGTGAAAGGAGGGAATCTCTGCTTGAGGATAATCCAGCT 138027
      :|||||
QY      163 IlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAspThr 182
      :|||||
Db      138028 GAGTTCTGTTAAATTTTATTTGGATGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 138087
      :|||||
QY      183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202
      :|||||
Db      138088 CTCATTCAGAGATTACAAAGTCCAGACCATCGATGAGTCAATGCTCCGGCTGAAGAA 138147
      :|||||
QY      203 GlyAspGlyGlyAsnAlaLysLysArgLys----- 212
      :|||
Db      138148 ATCAATGAAGCCAGAAAGAGCATAGAAAGTTTGGTGAGATTATGGGTTTCATCTTCAAAA 138207
      :|||||
QY      213 -----LysAlaSerAspAspValThrAspProArgValThrAsp----- 225
      :|||||
Db      138208 GTTGTATCTGACAAATTAACGATGATGATCCCTGATATTTGTACTACATCTGGCCCT 138267
      :|||||
QY      226 -----ProProValAspAspAspGluArgLysAlaSerGly----- 237
      :|||||
Db      138268 GCTCTTCAACCACTTGTATGTCACGACGACAAAA---TCAGGTATCACTATTATTAATAAA 138324
      :|||||
QY      237 ----- 237
      :|||||
Db      138325 TAATAATATTTTCATGTGCTGTGAATATGCTATAGCATTTTGTTCCTTTTCTTTTG 138384
      :|||||
QY      238 LysAspHisMetGly-----AlaLeuGluSerProLysVal 249
      :|||||
Db      138385 CGGCTGCACATTCGCACTTTTCTTAACACAGTTCAATCTGAAGGGTTCGGAAGCCCTGATGAG 138444
      :|||||
QY      250 IleThrLeuGlnSerSerCys----- 256
      :|||||
Db      138445 TACACCAAGATCTTCATGCCCAATGAAGCAAGAGTTGTCGNAATTAATTAAGGTGTT 138504
      :|||||
QY      257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsn----- 268
      :|||||
Db      138505 ACAACTACCGGTACGGATCATCTTAAGCTCTGTCAATTTTTTTCCTTATGATACACTTTA 138564
      :|||||
QY      268 ----- 268
      :|||||
Db      138565 GCATCATGCGCTGAAGTCATATCTATCTATCATATCTTTTTCGAAGCATGATGTGTACTA 138624
      :|||||
QY      269 -----AspAlaPheGly----- 272
      :|||||
Db      138625 GTTAGGTGCTTATGCATTGTGATGGCTACGGGTTTCAANATAACAGAAAATAACATACA 138684
      :|||||
QY      273 -----LeuTyrSerMet----- 276
      :|||||

```

```
Db 138685 TTATCCTATTACATGGTGAATGACTATATCAATATATATTTGAAGATACATACAGTTT 138744
Qy 277 ---GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp----- 293
Db 138745 AATGGAGCGAAATAGCAGGA-----CAAGTTCTCAATGA-AGTAGTGAGTGGATGTTG 138797
Qy 293 ----- 293
Db 138798 TTGTTTCCTACTAACTATCATTTTCCCAAGAGTGCGAAATAATAGATTGTGGAATCT 138857
Qy 293 ----- 293
Db 138858 TTCTTTGATCTTTACTAAAAATTTGAAGTTTCATGATGTGTGAGGTACACTCATCCAAGCT 138917
Qy 293 ----- 293
Db 138918 TAAGAGTAGCATCAATCATATAAAGAGTAGCGCTATCAACCTAGATTAAAAAGCAGCGATT 138977
Qy 293 ----- 293
Db 138978 CTCCACAACCTAACTCTTAAGATATTCACCATGAAAAAATTTGTAGAGGTGCAGTTATTC 139037
Qy 294 -----GlyLysIleProAspGluSerGlnArgArg 303
Db 139038 ATGGACTTAGCTCACTTAGCTATAGTGAACAAAAGCCCTCAAACTCAGTCAAGAGA 139097
Qy 304 Leu----- 304
Db 139098 TTACGGAGGCAAAAGTCCACTGTAAGAGAGCTCTAACTATTAAACACCTTGAATCATGAGA 139157
Qy 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis----- 320
Db 139158 CATTAGGGGAAAAAACCCTGAAGTAGAAAAGCTGGCTGT-GTCTGACACTCCAGCTGTCCG 139216
Qy 320 ----- 320
Db 139217 CAGTTTCAGTCACACTTGCTCTTAATGGGAAAAAATAATCTTTTAAGCTTTTTTTTGG 139276
Qy 320 ----- 320
Db 139277 TTTTATCATATAACTCTTACCTTTGTTAAACGCTGATCTCATTTATTGTTAAATTTTCTGTT 139336
Qy 321 -----LeuSerAspAlaCys 325
Db 139337 GCCTTTGGCTTGTCATCACCACTAATGCTTTTCGCATTAATTAATTAATTCAGGACAATGTT 139396
Qy 326 ThrSerMetValGlyAenPheLeuGluTrpValIleGluAenHisArgIleTyxGluGlu 345
Db 139397 ACCTTGGCTGGTGGAGCAGTCTTTTGAATATCTTTTGAATATCATGTAGTTGTTTCAGGAG 139456
Qy 346 ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla----- 361
Db 139457 CCAAAGTACATATTTCATGCGTTGACATAGCTTTGTTGACGAATACCAACCCCAATATC 139516
Qy 362 -----Ala-LeuLeuValLysGlnIle 368
Db 139517 GTCGTTCTTTTAATTTTCAGATTTCATATTGACCAAAATGCTAATGCTCGTAAATATCTTT 139576
Qy 368 eLeu-----SerHisLysGluSerLe 375
Db 139577 CCTTCAGTTGGGAGTGGCTCTATTTCATACTTTTAAGTGGATCATAAAGAGTCACT 139636
Qy 375 uValArgAlaAenSerGluLeuAlaPheLysCysSerArgValGluValAspTrpIleTy 395
Db 139637 TGCCCTTGCTGAAAAAGTTAAATACGATGCAATGAGGAGCTGGCAAGTGGTCTA 139696
Qy 395 rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs 415
Db 139697 TGACAGCCCTA-----AAGAGAAAAATTTCCAAAGAAAGCAGGTGCAACTGGCAGCAA 139747
Qy 415 pCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAenGluSerLe 435
Db 139748 TTGTCAATCAACA-----TCAGTAGAAAAGACCAAACTTCACAGCAGGAGAC 139795
```

```
Qy 435 uSerGlyAlaThrValArgAspGluLysIleAenThrLys----- 448
Db 139796 CTCTAAT---ATCTTGAGAAATGACCACATTTTCCCTAAGCAGAGAATGGATCTTTCATGA 139852
Qy 449 -----SerMetArgAenSerSerG1 455
Db 139853 TAATTTTCATGAATGGTCACCTTCAAGAGGCTCATTTGCTGCACAAATGGTTTTCAGA 139912
Qy 455 uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyxSerThrAl 471
Db 139913 AGAGCAGGAATGATAGCTGTGTCAGGAACCTCACATGGAATGC---CATTTTCAACTGA 139969
Qy 471 aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysTyxLysGlu 489
Db 139970 TGAATCTTCTGACATTTGCGAGAAAGAAATCAACTTAATTCACATGTTTCTCTCTGAAG 140029
Qy 489 nValGlnLysLeuValGlnGluHisGluLysMetLeuLysValGluAenMetTyxAl 509
Db 140030 AGAATACCGTATCTTTGATAAACAACAGCTCTCAGATTTTCAGAGTTTAGAGAATACACACA 140089
Qy 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAalaValIleArgIleTh 529
Db 140090 AAATAAACTGCCAGACTGAAACA-----GT 140116
Qy 529 rCysSerArgThrSerThrGlnValGlyAspLeuLysLeuAenPheHisAenTyxGluAr 549
Db 140117 GTGCAAT-----TTAGTTCTGGAAACATATTTGTAGAAG 140149
Qy 549 gLysPheAspGluIleLysSerGluLysAenGluCysLeuLysSerLeuGluGlnMetHi 569
Db 140150 TCATGCTGAT-----GTAGAGACCGAAATGACAAATAAAACAACTGTTGAGTGGTT 140203
Qy 569 sGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAenArgIleLysSe 589
Db 140204 CACTATGCTGATGATGATTCATTTCTGGAGCACATGAGACTCCAGCACACAACTTGAGTC 140263
Qy 589 r-----TrpAla-----AlaLysLeuLysValCysValProIl 600
Db 140264 ACTGCAATCTAATACATGGGCTGAAGACGACAGCTGAAGGAAAAAACTTTGTTGGAAGC 140323
Qy 600 eGlnSerGly----- 603
Db 140324 AAAATCAGGCCCAATATAGATCACACTTTTGATCAGCAGATTCCTTACACAGATTTCAAACCT 140383
Qy 604 -----AenAenLysHisPheSerGlySe 611
Db 140384 TGTTATGCAGAAATTCATCCATTTTGAAGGAACAAGTAGCAATTTCCCATGTCAGTGGAG 140443
Qy 611 rSer 612
Db 140444 CGCT 140447
RESULT 13
AP004752/c 156603 bp DNA linear PLN 21-JAN-2004
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
PAC clone:P0501G04.
ACCESSION
AP004752
VERSION
AP004752.3 GI:41052917
KEYWORDS
ORISAT
SOURCE
ORGANISM
Oriza sativa (japonica cultivar-group)
Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
REFERENCE
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
AUTHORS
Oriza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
TITLE
clone:P0501G04
JOURNAL
Published Only in Database (2002)
REFERENCE
2 (bases 1 to 156603)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
AUTHORS
```

TITLE Direct Submission
JOURNAL Submitted (13-FEB-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Jan 21, 2004 this sequence version replaced gi:34740248. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), POENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/db/glimmerm/glmr_form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/) -Splice predictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
 A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
 The orientation of the sequence is from SP6 to T7 of the PAC clone. This sequence of P0501G04 clone has an overlap with P0017C12 (DDBJ: AP004769) clone at 5' end and with QJ1399_H05 (DDBJ: AP004090) clone at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES
 source
 1..156603
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="P0501G04"
 complement(join(2875..3454,3805..3815))
 /gene="P0501G04.1"
 complement(join(2875..3454,3805..3815))
 /gene="P0501G04.1"
 /notes="gag-pol-like protein"
 complement(4646..7455)
 /gene="P0501G04.2"
 complement(join(4646..4662,5391..5429,6120..6360,7423..7455))
 /gene="P0501G04.2"
 /notes="start and end point are not identified"
 complement(join(4646..4662,5391..5429,6120..6360,7423..7455))
 /gene="P0501G04.2"
 /notes="predicted by RiceHMM etc."
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAD07829.1"
 /db_xref="GI:41052918"
 /translation="MKRKVSHSKSGSDVDKENVFKVPIERSFLDMLSSVGADLTQDKNKINELLIQHFGPDENCIDERAKNLLVDVFLVLSNSKPIVENTTVVDGIMKKLSKPDYVSA"
 10034..10234
 /gene="P0501G04.3"

gene
 misc_feature
 gene
 gene
 mRNA
 CDS
 CDS
 CDS

misc_feature

10034..10234
 /gene="P0501G04.3"
 /note="hypothetical ORF
 Predicted by GlimmerM
 this category is not included in IRGSP standard"
 complement(13016..16261)
 /gene="P0501G04.4"
 complement(join(13016..13562,14371..15279,15556..16261))
 /gene="P0501G04.4"
 /notes="supported by full-length cDNA(s): AK069014"
 complement(join(13328..13562,14371..15279,15556..16211))
 /gene="P0501G04.4"
 /notes="contains EST(s): AU031908 (R2510), D28321 (R2510)
 contains full-length cDNA(s): AK069014"
 /codon_start=1
 /product="kinesin light chain-like protein"
 /protein_id="BAD07830.1"
 /db_xref="GI:41052919"
 /translation="MPGLAANDSPVAAAPPPRRLLSPRLSPRSPAPSPSPSTSSRAKP
 RYVAAQPEDEALNDPDLGPFLKQADAMVSGEGGAARALEFAERAAARALERGE
 GABLELMSLHVAAATHCGLRHADAIFVLERAVAVYTPPPPEGESSAREQEP
 DQKGSEWALAFPGMMQLGDRHMRDESIACYGKLEIQMALGERDPRVATEC
 RYLAEAHQALQFDEAEKLCRKALETHREHSAPASLEASDRMLALIDAKDYDGA
 LHLVLASMTVMANGRDIEVATIDVAGTILALAFDEAVFSYQALTVLKGARDD
 HPSVASFVRLADLVHRTQRLRESKSCENALRVYAKPAPGAAPDEVAGDLMETAIY
 EALGDLEALKLQALKILESPGOMSTVAGIEAQMGLVYVMGVYAGDSRNSFENAV
 AKLRGERSKSAFFGVLLNQMLGACVQLFKIDEAAQLFEARAVLQEGGASHPTLG
 VYSNLAATIDAMGRVEDALELLEHLVKREKLGITANPDVEDEKLRLAELLKAGRSR
 NRKQKLENLFTVNSQRVKDAGRWSNFGFRS"
 18235..22976
 /gene="P0501G04.5"
 join(18235..18636,19319..19528,19610..20071,20165..20275,
 20362..20541,21136..21420,22634..22754,22855..22976)
 /gene="P0501G04.5"
 /notes="start and end point are not identified"
 join(18235..18636,19319..19528,19610..20071,20165..20275,
 20362..20541,21136..21420,22634..22754,22855..22976)
 /gene="P0501G04.5"
 /note="similar to Arabidopsis thaliana chromosome 1,
 FIK23.23"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAD07831.1"
 /db_xref="GI:41052920"
 /translation="MPSPSILLHHHHRLSPRAPSPSPRTLSLHPAPRLRLRL
 RSARGGSTRDIDGARSIFPARCYEARALRLDLSGTARREQAVAAALAAAGGAA
 EAHAAVADAMVNEAFDPGPGARRASMLRLILOANEVTEKASKUKKOIGTFFSEN
 EPDAESVLAMAFQVVMQDITNPRLEVPFSGSEIDQLSKPKRPADFSIRSDKL
 LAALABATFSCVIDARNNVLGGTGLFHKKNSCLSDSVCIDIISSESVNARSRR
 LQSFDLVQSHVAGKAKGWPAPKSERLVKIGPDPMWLSFVYTKLQIDAKAFK
 NTKLGGHVLANNKGVLSHAQMVLANILDMYFSDQTLPGKTFVSNWNSRPSKIK
 KNGYLNNLFPALLAGSSIVFLVGAIAQLCPQSLKRLKRLATVSSCVSSQSCSDIHS
 LQSSLEQAYCVSVSEKIKDSFGCTGDLMDANIGAVWGLPEFFKGINCDSDHYVDI
 QSTGAISQGEPSVSPISKMSHLQNDTQETLQNIASFQVVMSEKGVGFQPTN
 RLAVNHWATNPLTKLLEYGRKISPAFLERLIRSRPAKVVPVVELLMSVNSPFSALVR
 PVQDPC"
 complement(23753..26108)
 /gene="P0501G04.6"
 complement(join(23753..24276,24367..24619,24688..24918,
 25218..25348,25506..26108))
 /gene="P0501G04.6"
 /notes="supported by full-length cDNA(s): AK106946"
 complement(join(24086..24276,24367..24619,24688..24918,
 25218..25348,25506..26061))
 /gene="P0501G04.6"
 /notes="contains full-length cDNA(s): AK106946"
 /codon_start=1
 /product="putative GDSL-motif lipase/hydrolase protein"
 /protein_id="BAD07832.1"
 /db_xref="GI:41052921"
 /translation="MRRLLAVAAALLLVVHWFSLITMAAQLSPQPPDQYDDPPMP
 GLTVSPSPGYSDESPPLPDPSPPEPTPPQPPQPPSPPLPPRRRBPAP
 PLTVVPPQSPGWSVPPPARVINYYTTGCTTMLVFGDSTVDPCNNRLQTMKANFL


```
QY 294 -----GlyIysIleProAspGluSerGlnArg 303
Db 57773 ATGCACTTAGCTCATTATGCTATAGTGAAACAAAGCCCTCAAACTCAGTCAAGAGA 57714
QY 304 Leu-----
Db 57713 TTACGGAGGCAAAAGTCCACTGTAAGAGAGCTCTAACTATTAAACACCTTGAATCATGAGA 57654
QY 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis----- 320
Db 57653 CATTAGGGGAAAAAACCTGAAGTAGAAAGCTGGCTGT-GTCTGACACTCCAGCTGCCG 57595
QY 320 ----- 320
Db 57594 CAGTTCAGTCACACTTCTCTTAATGGGAAAAATAATTTCTGTAAAGTCTTTTTTTTGG 57535
QY 320 ----- 320
Db 57534 TTTTATCATATACTCTTACCTTTGTTTAACGCTGATCTCATTTATTTTCTGTT 57475
QY 321 -----LeuSerAspAlaCys 325
Db 57474 GCTTTGGCTTTGTCATCACCACTAAATGCTTTGGCAATTAATTAATTTGTCAGCAATGTT 57415
QY 326 ThrSerMetValGlyAenPheLeuGluTyrValIleGluAenHisArgIleTyrGluGlu 345
Db 57414 ACGTGTGCTGTGAGCAGTTCCTTTGAATATCTTTTGAATATCATATGATGTGTTTCAGGAG 57355
QY 346 ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla----- 361
Db 57354 CCAAGTACATATTTTCATGCGTTGAACATAGCTTTGGTACGAATAACCAACCCCAATATC 57295
QY 362 -----Ala-LeuLeuValLysGlnIle 368
Db 57294 GTCGTTCTTTTAATTTTCAGATTCATATGACCAAAATGCTAATGCTCGTAATAATCTTT 57235
QY 368 eLeu-----SerHisLysGluSerLe 375
Db 57234 CCTTCAGTGTGGGAGTGGCTCTATTATCACTTTAAGGTGGATCATTAAGAGTCACY 57175
QY 375 uValArgAlaAenSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTy 395
Db 57174 TGCCCTTGCTGAAAAAGGTTAAAAATACGAGTGCAATGAGGAGCTGGCAAGTTGGTCTA 57115
QY 395 rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs 415
Db 57114 TGACAGCCTA-----AAGAGAAAAATTTCCAAAGAAAAAGCAGGTGCAAACTGGCAGCAA 57064
QY 415 pCysPheGlyThrAenSerLysGlnSerValValSerThrLysLeuValAenGluSerLe 435
Db 57063 TTGTCAATCAACA-----TCAGTAGAAAAGACCAACCTTCACAGCAGGAGAC 57016
QY 435 uSerGlyAlaThrValArgAspGluLysIleAenThrLys----- 448
Db 57015 CTCTAAT--ATCTTGAGAAATGACCACATTTCCCTTAAGCAGAGAAATGATCTTCATGA 56959
QY 449 -----SerMetArgAenSerSerG 455
Db 56958 TAATTTTCATGAATGGTGCACTTCAAGAAGGCTCATTTGTTGCTGCACAAAATGGTTTCAGA 56899
QY 455 uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyrSerThrAl 471
Db 56898 AGAGCAGGAAGCTGATGCTGTTCCAGGAATCTCATGTGATGC---CATTTTCAACTGA 56842
QY 471 aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysLysTyrLysG 489
Db 56841 TGAATTCCTGACATTTGTGGAGAAAAAGAACTTAATTAATTGACAAATGTTTCTCTGTAAG 56782
QY 489 nValGlnLysLeuValGlnGluHisGluGluLysMetGluLeuLeuAenMetTyrAl 509
Db 56781 AGAATACCGTATCTTTGTATAACCAACAGTCTCAGATTTTCAGAGTTAGAGAAATACACACA 56722
```

```
QY 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaValIleArgIleTh 529
Db 56721 AATAAAACCTGCCAGACTGAAACA-----GT 56695
QY 529 rCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAenTyrGluAr 549
Db 56694 GTGCAAT-----TTAGTTCTGGAACATATTTGTAGAAG 56662
QY 549 GlySerPheAspGluIleLysSerGluLysAenGluCysLeuLysSerLeuGlnMethI 569
Db 56661 TCATGCTGAT-----GTAGAGACCCGAAATGACACAATAAAAAACAACCTGTTCAGTGGTT 56608
QY 569 sGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAenArgIleLysSe 589
Db 56607 CACTATGCTGATGATGATGCTATTTCTGGAGCAGCATGAGACTCCAGCAGCAAACTTGAGTC 56548
QY 589 r-----TrpAla-----AlaLysLeuLysValCysValProIl 600
Db 56547 ACTGCAATCTAATACATGCGCTGAGAGCGACAGCTGAAGGAAAAAACTTTGTTTGAAGC 56488
QY 600 eGlnSerGly----- 603
Db 56487 AAAATCAGGCCAATTAGATACACACTTTGATCAGCAGATGCTCTACCAAGATTCAAACTT 56428
QY 604 -----AsnAsnLysHisPheSerGlySe 611
Db 56427 TGTATTGCAAGAATTCATCCATTTGAAGGAACAAGTAGCAATTCCTCATGTCAGTGGAG 56368
QY 611 rSer 612
Db 56367 CGCT 56364
RESULT 14
LOCUS AR447512 5837 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 75 from patent US 6673549.
ACCESSION AR447512
VERSION AR447512.1 GI:42675836
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5837)
AUTHORS Furness,L.M. and Buchbinder,J.L.
TITLE Genes expressed in C3A liver cell cultures treated with steroids
JOURNAL Patent: US 6673549-A 75 06-JAN-2004;
FEATURES
source Location/Qualifiers
1..5837
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0.0041 Length: 5837
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 6 Gaps: 37
US-10-018-929C-3_COPY_650_1270 (1-621) x AR447512 (1-5837)
QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisAspValLysHisVal 21
Db 2760 GCGGACACAGTCTGTCATCTTGACTCTGACTGGAACCCCGAATGACTTGACGGCAAA 2819
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2820 GCCCGAGCGCATAGATTGGTCAGAGAAGCAGGTAATATTATTCGCTTAGTTACAAG 2879
QY 42 CysThrValGluGluLysAla-----Leu 49
Db 2880 GGGACTGTGGAGGAGGAGATCATAGAACGGGCCCAAAAGAGATGGTATTAGATCATCTG 2939
```

Qy	50	IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----	64
Db	2940	GTGATTAGCGCATGGACACCACTGGCGGACGATCTCGAAAAACAACCTCAGGAAGGTCC	2999
Qy	65	-----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
Db	3000	AACTCAAATCCTTTTAATAAAGAGAGACTGACACTATTTTGAATAATTTGGAGCAGAGAT	3059
Qy	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
Db	3060	CTCTTCAAAGAACTGGAA-----GGGAGGAATCA	3089
Qy	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
Db	3090	GAACCTCAGGAATGGAT-----ATAGATGAAATTTTGGCGTGGCTGAA	3134
Qy	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140
Db	3135	ACGAGAGAGATGAAGTGTCAACAAGTGCACAGATCAACTCTA-----TCCAGATTTTAAG	3191
Qy	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
Db	3192	GTTGCCAACTTTGCACACATGGAAGATGAAGAAGACTA-----GAAGACGCT	3239
Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer	180
Db	3240	CCTCAAGGACTGGGTGAGATCATT-----	3266
Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----	194
Db	3267	-----CCAGAGGAACAAGGAAAAAAGTAGAGGAGGAGCGCGCAGAGGAGCTAGAA	3320
Qy	195	-----GlySerGluAlaSerProLysThrGly	203
Db	3321	GAATTTATATGTCGCTCGAATTGCGAGTTCCTAAAGGCTTACAGACAATGACAGT	3380
Qy	204	AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal	223
Db	3381	GACTCTGACACTGAGTCTAAGAGCGAGCCAGAGATCCTCTGCTTCTGAGAGTGAAACG	3440
Qy	224	ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla	243
Db	3441	GAAGAC-----TCTGATGATGACAAGAAG-----	3464
Qy	244	LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly	263
Db	3465	-----CCAAAGCGCAGAGCGCTCCGAGGAGTGTCCGGAAGGACCTCGTGGAGGGA	3515
Qy	264	ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer	278
Db	3516	TTTACTGATGCAGAGATCCGAAGGTTTCATCAGGCTTATAAAGATTT-----	3563
Qy	279	HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle	296
Db	3564	-----GGTCTCCCTCTTGAACGCGCTGGAGTGCATAGCAGTGAT---GCTGAGCTG	3611
Qy	297	ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys	316
Db	3612	GTAGATAAGTCGGTGGCAGATCTG-----AAGCCGCTGGGT	3647
Qy	317	GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal	336
Db	3648	GAACGTGATCCAC-----AACAGCTGTGTGCACCAATGCAGGAATACGAGAGCAGCTG	3701
Qy	337	IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla	356
Db	3702	AAAGAAAATGCCAGCGAGGGAAGAGGACCGGAAAAAGG-----AGAGGTCCAAACAATCAAG	3758
Qy	357	LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal	376
Db	3759	ATATCCGAGATT---CAGGTTAATGTGAAATCCATTATCCAAACATGAAGAGGAGTTT---	3812

QY	377	ArgAla	AsnSerGlnLeuLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle	394
DB	3813	-----	-----GAGATGCTGCATAAATCTATCCCTGTGACCCCTGAAGAAAAA	3860
QY	395	TyrSer	IleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe	414
DB	3861	TACTGCTTAA	CTGCTGTCTGTCAAAGCTGCACATTTTGATGTAGAGTGGGGGTGGAGAT	3920
QY	415	AspCys	-----PheGly-ThrAsnSer-----Ly	422
DB	3921	GATTCTCGCCTGT	TGCTGGGATTATGAACATGGCTATGGAACCTGGGAGTTAATTAA	3980
QY	422	sGlnSer	ValValSerThrLysLeuValAsn-----	432
DB	3981	ACAG-----	-----ACCCAGAGCTTAAATTAACATGACAAAAATCTGCGGTGGAGACA	4028
QY	433	-----GluSer	LeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet	450
DB	4029	GATAAAGGCT	CAGGGAGACGCTACAGCCGAGCGGATTACTTTGTAAGCTGCTC	4088
QY	451	ArgAsnSer	SerGlu-----AspGluCysMetThrGlnLys	463
DB	4089	AGAAAGGGTCT	GGAGAGAGGGGCTGTGACAGTGGGGAGAGGCCCAATTAAGAAG	4148
QY	464	Arg-----	-----	464
DB	4149	CGGAAGCCTCGG	TAAAGAGAAAAACAAAGTGCCAGGCTGAAGAGGAGCATGGAATT	4208
QY	465	-----CysSer	HisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
DB	4209	GAGCTTTCATCT	CCTAGGCATTCAGATAATCCATCAGAGAGGGAAGTGAAGATGAT	4268
QY	481	GlyIleLysLys	LysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
DB	4269	GCCTTGGAAAA	AGTCCATGAAAAAACAAGAG--AAGAAAGAGAACAGGAGAAC	4325
QY	501	LysMetGlu	LeuLeuAsnMetTyrAlaAspLys-----	511
DB	4326	AAGGAGAAACA	ATAGTTCTAGGAAAGACAAAGAGGGGACAAGGAAAGAAAGAAAGTCA	4385
QY	512	LysGlnLys	LeuGluThrSerLysSerValGluAla-----	523
DB	4386	AAAGATAAGA	AAAGAGAGCCTTAAAGTGGTGATGTCCTCAAAATCTTCGAGTAAATCAAGCGA	4445
QY	524	-----AlaVal	IleArgIleTyrCysSerArgThrSerThrGlnValGlyAsp-----	539
DB	4446	TCTCAGGGTCT	GTGCCATATTACAGCAGGAGTGAACCTGTCCCCATTCGAGGAGTACG	4505
QY	539	-----	-----	539
DB	4506	GATGATGATCT	GGACCAGGAGACATTCAGCATATGTAAGGAGAGGATGATGCGCCGTGAAA	4565
QY	540	-----LeuLys	IleLeuAspHis-----AsnTyrGluArgLysPheAsp	552
DB	4566	AAGGCACTGAA	ACAGCTCGCAAAACCTCGACAAGGGGCTCAACCGTGAAGAACACAGCTGGAA	4625
QY	553	-----	-----GluIleLysSerGluLysAsnGluCysLeuLysSer---	564
DB	4626	CACACCGGAAC	TGCTGCTGAAAATTCGGAGACCGGATAGCCGATGCTTTAAGCCTAC	4685
QY	564	-----	-----	564
DB	4686	TCAGATCAGG	ACACATCAAACTCTGGAGGAGAACCTATGATTTTTTGTTCCTCAAGTTT	4745
QY	565	-----	-----LeuGlnMetHisGluValAlaLysLysLysLeuAla	577
DB	4746	ACAGATTTGAT	GCTCGAAAACTGCATAAGTTATACAAGATGGCTCATAGAAGAAAGTCT	4805
QY	578	GluAsp	GluAlaCysThrIleAsnArgIleLysSerTyrAlaAlaLysLeuLysValCys	597
DB	4806	CAAGAAGAGAG	-----GAGCAAAAGAAAGAAAGACAC	4838
QY	598	ValPro	IleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn	613

```
Db 4839 GTG-----ACTGGGGTGAAGAACCACTTCGTCAGAGGGCTCAGGCTCCAGCGG 4869
Qy 614 -----IleSerGlnAsn 617
Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 15
CO414477
LOCUS CO414477 7764 bp DNA linear PAT 23-JAN-2004
DEFINITION Sequence 21548 from Patent WO0170979.
ACCESSION CO414477
VERSION CO414477.1 GI:41322258
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lee,J. and Lillie,J.
TITLE Genes, compositions, kits, and method for identification,
assessment, prevention, and therapy of ovarian cancer
JOURNAL Patent: WO 0170979-A 21548 27-SEP-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..7764
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.00607 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 6 Gaps: 37

US-10-018-929c-3_COPY_650_1270 (1-621) x CO414477 (1-7764)
Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
Db 3342 GCGGACACAGTCGTCATCTTTGACTCTGACTGGAAACCCAGAAATGACTTGCAGGCACAA 3401
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3402 GCCCGAGCGCATAGAAATGGTGCAGAGAGCAGGTAATAATTTACCGCTTAGTTACAAAG 3461
Qy 42 CysThrValGluGluLysAla-----Leu 49
Db 3462 GGAGCTGTGGAGGGAGATCATAGACGGCCMAAAGAGATGGTATTAGATCATCTG 3521
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
Db 3522 GTGATTGAGCGCATGGACACCACTGGCGGAGCATCTGGAAACAACTCAGGAAGTCC 3581
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3582 AACTCAAAATCCTTTTAATAAAGAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGGAT 3641
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3642 CTCTTCAAGAACATGGAA-----GGGAGGAATCA 3671
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3672 GAACCTCAGGAAATGGAT-----ATAGATGAAATTTTTCGGTTGGCTGAA 3716
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3717 ACGAGAGAGAAATGAAGTGTCAACAAGTGCAACAGATGAACCTCTA-----TCACAGTTTAAAG 3773
```

```
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3774 GTTGCCCACTTTGCAACAATGGAAGATGAAGAGAGCTA-----GAAGACGCT 3821
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer 180
Db 3822 CCTCACAAGGACTGGGATGAGATCAATT-----3848
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
Db 3849 -----CCAGAGGAACAAGGAAAAAGTAGAGGAGGAGGAGCGCGAGAGGAGCTAGAA 3902
Qy 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3903 GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAAGGCTCAGACAAATGACAGT 3962
Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db 3963 GACTCTGACACTGAGTCTAAGAGCGCAGCCAGAGATCCTCTCTCTGAGAGTGAAACG 4022
Qy 224 ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 4023 GAAGAC-----TCTGATGATGACAAGAAG-----4046
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 4047 -----CCAAAGCCGACAGGGCGTCCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 4097
Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 4098 TTTACTGATGCAGAGATCCGAAGGTTTCATCAGGCTTATAAGAAGTTT-----4145
Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db 4146 -----GGTCTCCCTCTTGAACGGCTGAGTGCATACACGCTGAT---GCTGAGCTG 4193
Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4194 GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 4229
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 4230 GAACCTGATCCAC-----AACAGCTGTGTGTCAGCAATGCAGGAATACGAAGAGCAGCTG 4283
Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
Db 4284 AAGAATAATGCCAGCGAGGAAAGGACCGAGGAAAGG---AGAGTCCACAATCAAG 4340
Qy 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 4341 ATATCCGGAGTT---CAGGTTAATGTGAATCCATTATCCACATGAAGAGGAGTTT---4394
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db 4395 -----GAGATGCTGCATAAATCTATCCTCTGTGACCTCGAAGAAAAA 4442
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 4443 TACTGCTTAACCTGTGCTGTCAAAAGCTGCACATTTTGTATGTAGAGTGGGGGGTGAAGAT 4502
Qy 415 AspCys-----PheGly-ThrAsnSer-----LY 422
Db 4503 GATTCCTCGCTGTGTGGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATATA 4562
Qy 422 sGlnSerValValSerThrLysLeuValAsn-----432
Db 4563 ACAG-----ACCAGAGCTTAATAATTAACTGACAAAAATTTCTCGCGTGGAGACA 4610
Qy 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db 4611 GATAAAAGCCTCAGGGAAGCAGCTACAGCCGAGCGGATTTACTTTGTTGAAGCTGCTC 4670
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 09:34:25 ; Search time 865 Seconds
(without alignments)
4249.895 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILGSLNPSHDVKH.....SGNNKHFSGSSNTSQNAPDV 621

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/ROBINSON929/runat_11072005_102151_29159/app_query.fasta_1.2325
-DB=N_Geneseq_16Dec04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cgi
-LIST=45 -DOCALIGN=200 -THRT=1 -SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=ROBINSON929 @CGN 1 1 1234 @runat_11072005_102151_29159 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_16Dec04:*

- 1: geneseqn1980s:*
- 2: geneseqn1980s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3216	100.0	6571	4 AAA89354	Arabidops
2	2980	92.7	10329	4 AAA89353	Arabidops
3	2971	92.4	10329	4 AAA89385	Arabidops
4	494.5	15.4	467	4 AAA89379	Brassica
5	472	14.7	515	4 AAA89381	Brassica

C	6	442	13.7	519	4	AAA89377	Aaa89377 Brassica
	7	166	5.2	5220	12	ADQ59166	Adq59166 MSI-H car
	8	166	5.2	5837	12	ADL12346	Adl12346 Human ste
	9	166	5.2	7764	5	ADL63336	Adl63336 Human ova
	10	166	5.2	7764	10	AAD54628	Aad54628 Human chr
	11	166	5.2	7764	12	ADK60458	Adk60458 Angiogene
	12	166	5.2	7764	12	ADK60759	Adk60759 Angiogene
	13	166	5.2	7764	12	ADP73081	Adp73081 Angiogene
	14	162.5	5.1	4769	12	ADQ64522	Adq64522 Novel hum
	15	161	5.0	3045	13	AD848467	Ad848467 Bacterial
	16	160	5.0	5858	2	AAQ10378	Aaq10378 Encodes h
	17	159.5	5.0	5883	12	ADQ87119	Adq87119 Human tum
	18	159.5	5.0	5883	13	ADQ84854	Adq84854 Human tum
	19	159.5	5.0	5883	13	ACN40924	Acn40924 Tumour-as
	20	159.5	5.0	6432	5	AAS65903	Aas65903 DNA encod
	21	159.5	5.0	7274	13	ADN60282	Adn60282 Human non
	22	159.5	5.0	7396	6	ABT08505	Abt08505 Human nov
	23	159.5	5.0	7396	12	ADO09868	Ado09868 Human NOV
	24	159.5	5.0	7516	4	AAK51987	Aak51987 Human pol
	25	158.5	4.9	6354	10	ADJ56352	Adj56352 Chicken c
	26	158	4.9	5175	2	AAH57404	Aah57404 Human ske
	27	158	4.9	5896	4	AAH57404	Aah57404 Human ske
	28	156	4.9	5857	6	ABL61968	Ab161968 Colon ade
	29	156	4.9	5857	10	ADB31363	Adb31363 Bicalutam
	30	156	4.9	5857	13	ADP56352	Adp56352 Human PRO
	31	156	4.9	5943	4	AAH72966	Aah72966 Human cer
	32	156	4.9	6028	4	AAH72966	Aah72966 Human cer
	33	156	4.9	6364	8	ACA04019	AcA04019 cDNA down
	34	155.5	4.8	6965	10	ADD29780	Add29780 Human tum
	35	155.5	4.8	7636	10	ACF58150	Acf58150 Human gol
	36	155.5	4.8	7694	6	ABX04175	Abx04175 Human mRN
	37	155.5	4.8	7694	8	ABX10347	Abx10347 DNA encod
	38	155.5	4.8	7695	6	ABL68595	Ab168595 Kidney ca
	39	155.5	4.8	7695	10	ACF58151	Acf58151 Human gol
	40	155.5	4.8	7743	10	ACF58152	Acf58152 Human gol
	41	155.5	4.8	7761	10	ACF58153	Acf58153 Human gol
	42	155.5	4.8	8063	5	ADL63564	Adl63564 Human ova
	43	149	4.6	5847	6	ABZ32324	Abz32324 Candida a
	44	148	4.6	4368	12	ADQ59186	Adq59186 MSI-H car
	45	147.5	4.6	4944	6	ABZ32285	Abz32285 Candida a

ALIGNMENTS

RESULT 1	AAA89354	standard; CDNA; 6571 BP.
ID	AAA89354	
XX	AAA89354;	
AC	AAA89354;	
XX	23-APR-2001 (first entry)	
DT	23-APR-2001	
XX	Arabidopsis thaliana silencing gene cDNA.	
DE	Arabidopsis thaliana silencing gene cDNA.	
XX	Gene silencing; silencing gene; MOM; ss.	
KW	Gene silencing; silencing gene; MOM; ss.	
XX	Arabidopsis thaliana.	
OS	Arabidopsis thaliana.	
XX	Key Location/Qualifiers	
PH	CDS	310..6315
FT		/*tag= a
XX	WO200100801-A2.	
XX	04-JAN-2001.	
PD	04-JAN-2001.	
XX	21-JUN-2000; 2000WO-EP005761.	
PF	21-JUN-2000; 2000WO-EP005761.	
XX	23-JUN-1999; 99GB-00014623.	
PR	(NOVS) NOVARTIS AG.	
XX	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
XX	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.	

XX	21-JUN-2000; 2000WO-EP005761.	Db	4472	GAACATCTATTATGACGCGGTGATTTCATGAATCTCGTCCATCTTTCTTCCAAAGGT	4531
XX	23-JUN-1999; 99GB-00014623.	Qy	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140
XX	(NOVS) NOVARTIS AG.	Db	4532	GGAGAAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAT	4591
XX	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.	Qy	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisLeuLeuSerAspGluGluSer	160
XX	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;	Db	4592	TACAGCAGTGATTCTACTCTATTGTGGTGAAGACCATAATTAAAGTTGTAGATGAAGAGAGT	4651
XX	WPI; 2001-137952/14.	Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpProSer	180
XX	P-PSDB; AAB20062.	Db	4652	CCAAATATATTTGGTCAAAAGCTGTTGGGGGGAATAATCCTATGTGGAATAACCTTCA	4711
PT	Novel gene encoding a protein that controls gene silencing, in particular	Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
PT	silencing of plant genes.	Db	4712	GATACTCCCAAGGAATCGAAAACGAGTTTCAGTATTTTGGGGTTCGAAAGGAGTCCC	4771
PS	Claim 6; Page 19-22; 48pp; English.	Qy	201	LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp	220
XX	The present sequence is that of an Arabidopsis thaliana 'Gene (MOM)	Db	4772	AAACTGGCGATGGTGGAAATGCAAAAGAACGAAAGAGCTTCTGATGATGTCATGAT	4831
CC	involved in epigenetic gene silencing. It encodes a 2001-amino acid	Qy	221	ProArgValThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHis	240
CC	protein (see AAB20062) with similarity with Arpase/helicase proteins of	Db	4832	CCCGGGTCACTGATCCGCCAGTAGATGATGATGATGATGATGATGATGATGATGATGAT	4891
CC	the SWI2/SNF2 family, which controls gene silencing, particularly plant	Qy	241	Met-----	241
CC	gene silencing. Trans-acting modified loci were identified by T-DNA	Db	4892	ATGGGTAAATAGTATTTTCTGCTCCGATACCTCTAGTGTTCATTGATGATGATGATGAT	4951
CC	insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which	Qy	242	-----GlyAlaLeuGluSerProLysValIleThrLeuGln	253
CC	carries a heritable inactivated, methylated hygromycin resistance gene	Db	4952	CTTTGCTGACTATCTTTCTTACAGGGGCT--TTGAGTCAACAAAGTCAATACACTCCAG	5010
CC	(hyg). Mutation of a silencing modified locus results in release of	Qy	254	SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273
CC	silencing of the hyg gene and restores hygromycin resistance. Plants	Db	5011	TCATCATGTAAATCTTCTGGTACAGATGGTACATGATGATGATGATGATGATGATGATGATG	5070
CC	homologous for the silent resistance gene were subjected to	Qy	274	TyrSerMetGlySerHisLysSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293
CC	transformation with a bar selectable marker gene under control of the T-	Db	5071	TATTCATGGCAGCCATATCTCTGGAATCCCAAGAGATATGTTAGTCAAGATTGG	5130
CC	DNA 1'-2' dual promoter. Transformants were selected and their progeny	Qy	294	GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla	313
CC	screened for hygromycin resistance. The mutant phenotype (hygromycin	Db	5131	GGGAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTTAAAGCCGGAAGATGGCA	5190
CC	resistance) was screened for genetic co-segregation with a specific T-DNA	Qy	314	LysLeuCysGlnValLeuHisLeuSer-----	322
CC	insert. Cloning of the tagged gene allowed characterization of the	Db	5191	AAACTTTGCCAAGTTTTCATCTTC-AGTAAAGTGGCCTTTTTCACCTCCCAACTATT	5249
CC	present sequence. 2 Allelic differences were found in the corresponding	Qy	322	-----	322
CC	gene of A. thaliana ecotype Columbia, occurring at nucleotide positions	Db	5250	TTAGCCTTGCATATGCTTATATATAGTGTGATGCAACTGTAGTTGTTACCTGATTTCTGT	5309
CC	4338 (A instead of T) and 6721 (T instead of G). Gene silencing is useful	Qy	322	-----	322
CC	as a molecular tool for regulating gene expression	Db	5310	TTACAGCCAAATGTGAGAGTTTATTCTTCAACTATATCCATCCGTTTAAAGCATATTTTA	5369
CC	Revised record issued on 04-NOV-2004 : Correction to feature table key	Qy	322	-----	322
XX	Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;	Db	5370	TTTCTTATATCTGGCTTCGTTACCAATGCACCTGTAAATAGCAACTGTCTGCACAAAC	5429
SQ	Alignment Scores:	Qy	322	-----	322
	Pred. No.: 3,43e-275 Length: 10329	Db	5430	AGTAGGTAGTTATGTGCTCATGTCTTATTTGTTTATTGAAGCAAGAAATTTCTGTCT	5489
	Score: 2980.00 Matches: 618	Qy	322	-----	322
	Percent Similarity: 68.51% Conservative: 0	Db	5490	ACTTTACATGATCCATCTCTGGGAGTATATAACTATATATAACCTTAGGCTTTGTACCT	5549
	Best Local Similarity: 68.51% Mismatches: 3	Qy	322	-----	322
	Query Match: 92.66% Indels: 284	Db	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
	DB: 4 Gaps: 3	Qy			
	US-10-018-929C-3_COPY_650_1270 (1-621) x AAA89353 (1-10329)				
Qy	1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20				
Db	4172 CGAGCTGATCGGTTATTCTTTTGGGAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC 4231				
Qy	21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40				
Db	4232 GTAGAGAAGATAAAATCGAGTCATGTCTCTGAAGAACTAAGATATTCCGATTGTACTCA 4291				
Qy	41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60				
Db	4292 GTATGTACAGTTGAAGAAAAGCCCTGATTTCTGGCTAGGCCAAATATCGGGCAAAATPAA 4351				
Qy	61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80				
Db	4352 GCTGTAGAGAACCTAAACCGCTCTCTCACGACGACACTGCTCATGTGGGGGGCTCATAC 4411				
Qy	81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100				
Db	4412 TTTATTTGATAAATCGGATCATTTTTCACAGCAGTGAATCTCAGATTCAGGAGTTTCATT 4471				
Qy	101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120				

Db 5550 GGCTGATCAAGACATGTCAAAGTTTATCTGTTGCTGCTGGTATAGAACTAATACAG 5609
Qy 322 ----- 322
Db 5610 TGTCTGATGCTATTATTAAGGTAGTCTTATGCTTCCACATATTGGCTAATAGATGTTTCCG 5669
Qy 322 ----- 322
Db 5670 CTGCTGTCATATACCTTCTGTGATATACAGGTGCTCCGTCTATCAAAATTTGACTAA 5729
Qy 323 ----- AspAlaCysThr 326
Db 5730 AAGGTATTTCGAATGTGTGATTGGTTAAACAGATTATTTGTTTTCAGGATGCTTGACA 5789
Qy 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluLeuPro 346
Db 5790 AGCATGGTCGAAATTTCTCGAATATGTTATGAAATACCGAATCTACGAAGAGCCA 5849
Qy 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
Db 5850 GCCACTACTTTTCAGGCAATCCAGATAGCCCTT-GGTATGACAGCATTTACTTTGTATAATT 5908
Qy 357 ----- 357
Db 5909 TATGCATTGTTTCCTTCATCATCTGCCCTTTGTTAGAAATGTCCTCAGNAGCAGCACTCC 5968
Qy 357 ----- 357
Db 5969 TTTAGTTTAACTTTTCCAATCATAGATTCAAATATCCATTAACTGGCCTTTGATCGCTG 6028
Qy 358 -----SerTyr 359
Db 6029 CATAATATGAATAGTTGACATACTGAATACGTTGTTTAATATGCAATTTTCAGAGTTGG 6088
Qy 360 IleAlaIleLeuLeuValIleGlnIleLeuSerHisIleValSerLeuValArgAlaAsn 379
Db 6089 ATTGACGCTTGTGTTAAGCAATTTCTAGCACAAAGAAATCTTGCTCCGTCGAAT 6148
Qy 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
Db 6149 TCTGAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGGATTATATTTATTCGATATTGTCC 6208
Qy 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 6209 TGCATGAAGAGTCTGTTCTCGAGCATACACAAGGTTTGCAAGTTTCGATTGTTGGTACT 6268
Qy 420 AsnSerLysGlnSerValIserThrLysLeuValAsnGluSerLeuSerGlyValAlaThr 439
Db 6269 AATTCTAAACAGTCAGTGGTTAGCACAAAACCTAGTAATGAAGTCTCTCAGGGGCTACA 6328
Qy 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
Db 6329 GTGCGTGACGAAAGATTAAATACGAAGTCGATGCGAAATAGCTCAGAGGATGAAGATGC 6388
Qy 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
Db 6389 ATGACTGAGAAGAGATGTAGCCATTATAGCACAGCAACAGAGATATCGAAAGACTATT 6448
Qy 480 SerGlyIleLysLysLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
Db 6449 AGTGGCATAAAAAGAAATAACAAGCAAGTGAAGCTTTGTACAGAGCATGAGGAA 6508
Qy 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
Db 6509 AAGAAATGGAGCTGTTAAATATGTATGCACAGAAGCAGCAAACTTGAACCTAGTAAA 6568
Qy 520 SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
Db 6569 AGTGTGAAGCAGCAGTAATTCTGATTACCTGTTCCAGCAGCACTACTCAAGTGGGTGAT 6628
Qy 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn 559
Db 6629 CTCAAACTGCTGGATCATTAATTATGAAGAAAGTTTGTATGAATCAAAAGTGAGAAAAAT 6688

Qy 560 GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
Db 6689 GAATGCTCTCAAAAGCTTGGAGCAAAATGCAGGTTTGCAAAAGAAAGTTTGGCTGAGGAT 6748
Qy 580 GluAlaCysThrIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro 599
Db 6749 GAAGCCTGTTGGATTATCGGATAAGAGCTGGGCAGCTTAATTAAGAAGTTTGTGTTCCC 6808
Qy 600 IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro 619
Db 6809 ATTCAAGTGCATAACAAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCT 6868
Qy 620 AspVal 621
Db 6869 GATGTA 6874
RESULT 3
AAA89385
ID AAA89385 standard; DNA; 10329 BP.
XX
AC AAA89385;
XX
DT 04-NOV-2004 (revised)
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX KW Gene silencing; silencing gene; MOM; ds.
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT exon 1..1008
FT /*tag= a
FT /number= 1
FT intron 1009..1295
FT /*tag= b
FT /number= 1
FT exon 1296..2550
FT /*tag= c
FT /number= 2
FT CDS 1301..9808
FT /*tag= d
FT /product= "Columbia protein"
FT /note= "This sequence contains 15 introns."
FT intron 2551..2673
FT /*tag= e
FT /number= 2
FT exon 2674..2752
FT /*tag= f
FT /number= 3
FT intron 2753..2867
FT /*tag= g
FT /number= 3
FT exon 2868..3113
FT /*tag= h
FT /number= 4
FT intron 3114..3506
FT /*tag= i
FT /number= 4
FT exon 3507..3680
FT /*tag= j
FT /number= 5
FT intron 3681..3973
FT /*tag= k
FT /number= 5
FT exon 3974..4895
FT /*tag= l
FT /number= 6
FT allele 4338
FT /*tag= m
FT /note= "ecotype Zurich"

FT intron 4896.. .4975
FT /*tag= n
FT /number= 6
FT 4976.. .5217
FT /*tag= o
FT /number= 7
FT 5218.. .5777
FT /*tag= p
FT /number= 7
FT 5778.. .5882
FT /*tag= q
FT /number= 8
FT 5883.. .6082
FT /*tag= r
FT /number= 8
FT 6083.. .7480
FT /*tag= s
FT /number= 9
FT 6721
FT /*tag= t
FT /note= "ecotype Zurich"
FT 7481.. .7615
FT /*tag= u
FT /number= 9
FT 7616.. .7771
FT /*tag= v
FT /number= 10
FT 7772.. .7914
FT /*tag= w
FT /number= 10
FT 7915.. .8070
FT /*tag= x
FT /number= 11
FT 8071.. .8153
FT /*tag= y
FT /number= 11
FT 8154.. .8318
FT /*tag= z
FT /number= 12
FT 8319.. .8451
FT /*tag= aa
FT /number= 12
FT 8452.. .8629
FT /*tag= ab
FT /number= 13
FT 8630.. .8718
FT /*tag= ac
FT /number= 13
FT 8719.. .8918
FT /*tag= ad
FT /number= 14
FT 8919.. .9000
FT /*tag= ae
FT /number= 14
FT 9001.. .9211
FT /*tag= af
FT /number= 15
FT 9212.. .9284
FT /*tag= ag
FT /number= 15
FT 9285.. .10329
FT /*tag= ah
FT /number= 16
XX WO200100801-A2.
XX
XX
XX 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-EP005761.
XX
XX 23-JUN-1999; 99GB-00014623.
XX
XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20063.
XX
PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
XX Claim 5; Page; 48pp; English.
XX
CC The present sequence is that of an Arabidopsis thaliana ecotype Columbia
CC gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino
CC acid protein with similarity with Arpase/helicase proteins of the
CC SWI2/SNF2 family, and which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Zurich, occurring at nucleotide positions
CC 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful
CC as a molecular tool for regulating gene expression. Note: the present
CC sequence is not shown in the specification but is derived from the A.
CC thaliana mutant silencer gene sequence given in AAA89333
CC
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
SQ Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,52e-274 Length: 10329
Score: 2971.00 Matches: 616
Percent Similarity: 68.40% Conservative: 1
Best Local Similarity: 68.29% Mismatches: 4
Query Match: 92.38% Indels: 284
DB: Gaps: 3

US-10-018-929C-3_COPY_650_1270 (1-621) x AAA89385 (1-10329)

Qy 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
Db 4172 CGAGCTGATGCGTTTATCTCTTTTGGAGCAGCTTGATCCATCGCATGATGTTAGCAC 4231
Qy 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
Db 4232 GTAGAGAGATATAAATCGAGTCATGTTCTGAAGAAGCAATTAAGATATCCGATTGTACTCA 4291
Qy 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
Db 4292 GTATGTACAGTTGAAGAAAAAGCCCTGATTCTGGCTAGGCAAAATAAGCGGCAATAA 4351
Qy 61 AlavalGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 4352 GCTGTAGAGAACTTAACCGCTCTCTCACGACGCACTGCTCATGTGGGGGGGTCTATAC 4411
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 4412 TTATTTGATAAATCTGGATCATTTTTCACAGCAGTGAAATCCAGATTCCAGAGTTTCATT 4471
Qy 101 GluGlnSerIleMetAspGlyValIleHisGlnPheSerSerIleLeuSerSerLysGly 120
Db 4472 GAAACAATCTATTATGACGCGTGATGAAATTCGTCATATCTTCGTCATATCTTCGTCAT 4531
Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140

Db 4532 GGAGNAGAAATGAAGTCAAGCTGTCTACTCTTTTGGAGGCCAAGCATGCTCAGGGAAC 4591
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 4592 TACAGCAGTGAATCTACTCTAATTTGGTCAAGACCATATAAAGTTGTCAGATGAAGAGAGT 4651
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyProSer 180
Db 4652 CCAATATATATTTGGTCAAAAGCTGTGGGGGGAATAATCTCTATGTGGAAATACCCCTTCA 4711
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 4712 GATACTCCCCAAGGAATCGAAACGAGTTCAAGTATTTTGAAGGGTCTGAAAGCGAGTCCC 4771
Qy 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp 220
Db 4772 AAACTGGCGATGGTGGAAATGCAAAAGAGCAAGAAAGGCTTCTGTGATGATGTCACTGAT 4831
Qy 221 ProArgValThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 4832 CCCCAGGTCACATGATCCGCCAGTAGATGATGATGAAGAAAGGCTCTGGGAAGGATCAC 4891
Qy 241 Met----- 241
Db 4892 ATGGTAAATAGTTTAATTTCTGCTCCGATACCTCTAGTCTTCAATGATATGCAACTA 4951
Qy 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 4952 CTTTGTCTGACTATCTTCTCCTACAGGGCT-TTGGAGTCACCAAAAGTCATAAAGCTCCAG 5010
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyLysAsnAspAlaPheGlyLeu 273
Db 5011 TCATCATGTAAATCTCTGTGTACAGATGATGATGATGATGATGATGATGATGATGATGATG 5070
Qy 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlyAspTrp 293
Db 5071 TATTCATATGGCAGCCATATCTCTGGAAATCCAGAGGATATGTTAGCTAGTCAAGATTGG 5130
Qy 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
Db 5131 GGGAAATATACCGATGAATCACAGAGGAGGCTCCACACATGTTTTAAAGCCGAAGATGCA 5190
Qy 314 LysLeuCysGlnValLeuHisLeuSer----- 322
Db 5191 AAACCTTTGCCAAGTTTGTGCATCTTTC-AGTAAAGTGGCCTTTTTCACCTCCACAACTTAT 5249
Qy 322 ----- 322
Db 5250 TTAGCCTTGCATATGCTTATATATATAGTGAATGCAACTGTAGTGTGTACCTGATTTCTGTG 5309
Qy 322 ----- 322
Db 5310 TTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATATCCATCCGTTTAAGCATATTTA 5369
Qy 322 ----- 322
Db 5370 TTTCTTATATCTGGCTTCGTTTACCAATGCACTGTTAAAAATGAGCAACTGCTGCACAAAAAC 5429
Qy 322 ----- 322
Db 5430 AGTAGGTAGTTATGTGCTCAATGCAATTCATTTGTTTATTTGAAGCAAAAGAAATTTCTGTCT 5489
Qy 322 ----- 322
Db 5490 ACTTTACATGATCCTGTGGAGATATATAATATATATATAACCTTAGGCCCTTTGTACCT 5549
Qy 322 ----- 322
Db 5550 GGCTGATCAAAGACATGTCAAAGTTTATCTGTTCCGCTGTGGTATAGAAACTAATACAG 5609
Qy 322 ----- 322

Db 5610 TGTCTGATGCTATTTTAAAGGTAGTCTTATGCTTCACATATTGGCTAATAGATGTTTCCG 5669
Qy 322 ----- 322
Db 5670 CTGTGCTGCCATATACTTCTGTGATTATACGGGTGCTCCGCTCTATCAAAATTTGTACTAA 5729
Qy 323 -----AspAlaCysThr 326
Db 5730 AAGGTATTTTGCAAATGTGTGATTGGTTAACAGATTATTTTGTGTTTTCAGGATGCTTGCA 5789
Qy 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
Db 5790 AGCATGTGCGGAATTTTCTCGAATATGTTATTGGAANAATCACCCGATCTACGAAGAGCA 5849
Qy 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
Db 5850 GCCACTACTTTTTCAGGCATTCAGATAGCCCT-GGTATGACAGCATTTTACTTTTGATAATT 5908
Qy 357 ----- 357
Db 5909 TATGCATTTGTTTCTTCATCATCTGCCTTTGTTTGAATGTCCTCAGAAGAGCAGCACTCC 5968
Qy 357 ----- 357
Db 5969 TTTAGTTTAACTTTCCCAATCATAGGATTCAAAATATCCATTAACCTGGCCTTTGATCGCTG 6028
Qy 358 -----SerTrp 359
Db 6029 CATAATATATGAATAGTTGACATATCGAATACGTTGTTAATAATGCAATTTTCAGAGTTGG 6088
Qy 360 IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
Db 6089 ATTGCAGCCTTGTGTGTAAGCAAAATCTTAGCCACAAAGAAATCTCTGTCCGTGCAAAAT 6148
Qy 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
Db 6149 TCTGAAATAGCTTTCAAATGCTCTAGAGTAGAGGTGGAATATATTTATTCGATATTGTCC 6208
Qy 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 6209 TGCATGAAGAGTCTGTTCTCTGGAGCATACCAAGTTTGCAGTTTCGATTTGTTGGTACT 6268
Qy 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyValaThr 439
Db 6269 AATTCTAAACAGTCAGTGGTTAGCACAAAATCTAGTAAATGAAGTCTCTCAGGGGTACA 6328
Qy 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
Db 6329 GTGCGTCACGAAAGATTAAATACGAAGTCAATGCGAAATAGCTCAGAGGATGAAGAGTGC 6388
Qy 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
Db 6389 ATGACTGAGAAGAGATGTAGCCATTATAGCACACACACAGAGATATCGAAAGACTATT 6448
Qy 480 SerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
Db 6449 AGTGCCATAAAAAAGAAATAACAAGAGCAAGTGTCAAAAGCTTGTACAAGAGCATGAGAA 6508
Qy 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
Db 6509 AAGAAATGGAGCTGTTAAATATATGTATGACAGCAAGAGCAAGAACTTTGAACCTAGTAAA 6568
Qy 520 SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
Db 6569 AGTGTGGAAGCAGCAATTCGTATTTACCTGTTTTCACGGACCACTACTCAAGTGGTGAT 6628
Qy 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluLysSerGluLysAsn 559
Db 6629 CTCAAACTGTGATCATATAATTTGAAAGAAAGTTTGTGATAAATCAAAAGTGAAGAAAT 6688
Qy 560 GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLeuAlaGluAsp 579
Db 6689 GAATGCCTCAAAAGTCTGGAGCAAAATGCAGATGTTTGCAAAGAAAGAGTTTGGCTGAGGAT 6748

```

QY 580 GluAlaCysTrpIleAenArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro 599
Db 6749 GAAGCCGTGTGGATTAAATCGATAAAGAGCTGGCAGCTAAATTAAGTTGTGTCC 6808
QY 600 IleGlnSerGlyAenLysHisPheSerGlySerSerAenIleSerGlnAenAlaPro 619
Db 6809 ATTCAAAGTGGCAATAACAAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCT 6868
QY 620 AspVal 621
Db 6869 GATGTA 6874

RESULT 4
AAA89379
ID AAA89379 standard; DNA; 467 BP.
AC AAA89379;
XX
DT 11-SEP-2003 (revised)
DT 23-APR-2001 (first entry)
XX
XX Brassica oleracea silencing gene homologous DNA seq1-43.
XX DE
XX KW Gene silencing; silencing gene; MOM; ds.
XX OS Brassica oleracea; acephala.
XX PN WO200100801-A2.
XX PD 04-JAN-2001.
XX PF 21-JUN-2000; 2000WO-EP005761.
XX PR 23-JUN-1999; 99GB-00014623.
XX PA (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX DR WPI; 2001-137952/14.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
XX silencing of plant genes.
XX
XX Example 6; Page 45; 48pp; English.
XX
XX This is the sequence of Seq1-43, a genomic DNA sequence of Brassica
XX oleracea acephala that shows homology to the newly identified silencer
XX gene (see AAA89353) of Arabidopsis thaliana. Seq1-27 is a partial
XX sequence of clone 1, which was isolated from a genomic library screened
XX with MOM cDNA. Partial sequences of clone 1 (see also AAA89377-80) showed
XX similarity to different regions of the MOM gene (80-86% at DNA level and
XX 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-
XX terminal parts of the MOM protein (see AAB20062). Gene silencing is
XX useful as a molecular tool for regulating gene expression. (Updated on 11
XX -SEP-2003 to standardise OS field)
XX
XX Sequence 467 BP; 145 A; 87 C; 108 G; 127 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.76e-38 Length: 467
Score: 494.50 Matches: 97
Percent Similarity: 84.09% Conservative: 14
Best Local Similarity: 73.48% Mismatches: 12
Query Match: 15.38% Indels: 9
DB: 4 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x AAA89379 (1-467)

QY 111 GluPheSerSerIleLeuSerSerLysGlyGluAenGluValLysLeuCysLeu 130
|||

```

```

Db 1 GAATCTCGTCCATACATCTTCTCCGATGTTGGAGAAAGAAATGAAGCAAGCTGTGTCTA 60
QY 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu 150
Db 61 CTTTTCGAAGCCCAAGCATGCTCAGGAAGTTACAGCAGCTGATCTACTTATTTGGTGAA 120
QY 151 AspHisIleLysLeuSerAspGluGluSerProAenIlePheTrpSerLysLeuLeuGly 170
Db 121 GAACATGTCAGTTATCAGAT---GAAGTCCAAATATGTTTGGTCAAGAGCTGTGAGT 177
QY 171 GlyLysAenProMetTrpLysTyrProSerAspThrProGlnArgAenArgLysArgVal 190
Db 178 GGAAGAAGACCCATATGTGGAATACTGTCGGATACTCCTCAAAGGAGTCGAAAAAGAGTA 237
QY 191 GlnTyrPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAenAlaLysLys 210
Db 238 CGGCATCTTCAGGGCTATGAGGAGACTACCAAGTTGGCAATGGCGAAACTTAAAGAG 297
QY 211 ArgLysLysAlaSerAspValThrAspProArgValThrAspProProValAspAsp 230
Db 298 AAAAAGAGGCTTCAGATGATGTCA---GTAGATAAC 333
QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGly 242
Db 334 GCTGAGAGAAAGCCTCTCGAAAGGATCACATGGGT 369

RESULT 5
AAA89381
ID AAA89381 standard; DNA; 515 BP.
AC AAA89381;
XX
DT 11-SEP-2003 (revised)
DT 23-APR-2001 (first entry)
XX
XX Brassica oleracea silencing gene homologous DNA seq2-33.
XX DE
XX KW Gene silencing; silencing gene; MOM; ds.
XX OS Brassica oleracea; acephala.
XX PN WO200100801-A2.
XX PD 04-JAN-2001.
XX PF 21-JUN-2000; 2000WO-EP005761.
XX PR 23-JUN-1999; 99GB-00014623.
XX PA (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX DR WPI; 2001-137952/14.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
XX silencing of plant genes.
XX
XX Example 6; Page 46; 48pp; English.
XX
XX This is the sequence of Seq2-33, a genomic DNA sequence of Brassica
XX oleracea acephala that shows homology to the newly identified silencer
XX gene (see AAA89353) of Arabidopsis thaliana. Seq2-33 is a partial
XX sequence of clone 2, which was isolated from a genomic library screened
XX with MOM cDNA. Partial sequences of clone 2 (see also AAA89382-84) showed
XX similarity to different regions of the MOM gene (64-76% at DNA level and
XX 55-64% at amino acid level) which encode the ATPase, putative
XX transmembrane and C-terminal parts of the MOM protein (see AAB20062).
XX Gene silencing is useful as a molecular tool for regulating gene
XX expression. (Updated on 11-SEP-2003 to standardise OS field)
XX
XX Sequence 515 BP; 181 A; 74 C; 136 G; 124 T; 0 U; 0 Other;

```

Alignment Scores:

Pred. No.: 8.03e-36 Length: 515
 Score: 472.00 Matches: 100
 Percent Similarity: 60.87% Conservativity: 26
 Best Local Similarity: 48.31% Mismatches: 39
 Query Match: 14.68% Indels: 42
 DB: 4 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x AAA89381 (1-515)

QY 387 SerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeu 406
 DB 1 TCTAGAGAGGCTGGATATGATATCTTTCTGTACTGATGAAGAGCTATCTGTG 60
 QY 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
 DB 61 GGGCGCACACAAAGGTTTCCAA----- 81
 QY 427 SerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn 446
 DB 81 ----- 81
 QY 447 ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSer 466
 DB 82 -----GAAAGGGTGAAGAAATGCTGCTGAGAAAGAGGTAGC 120
 QY 467 HisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTyr 486
 DB 121 CATATAGCTCAGTACCAAGGATGTTGAAAGACTATTAGCGCATCAAAAAGAAATGC 180
 QY 487 LysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsn 506
 DB 181 AGTAAGAGCCTTCATAGCTGTACAAACCTCGAGGAAGAAAGATGCACTGATGAAT 240
 QY 507 MetTyrAlaAspLysGlnLysLeuGluThrSerLysSerValGluAlaValIle 526
 DB 241 AGGAATGCTGTCAAGACGAGAACTTCAGAAATTTGTAAGAGGTGAAGATCAATTTAT 300
 QY 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
 DB 301 CGTGTCACTATTACAGGTATAAATACTCAGAGCTTACATGATGCTCTCCACGGCTGAA 360
 QY 545 HisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSer 564
 DB 361 TGTACTTTTGAAGAAAGTTTGATGATCTCAAGAGGAGAGTGGATGAATGCTTGAAAGT 420
 QY 565 LeuGluGlnMetHisGluValAlaLysLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
 DB 421 TTAGAGCAAAATAACGAGGCTGGAAAGAAAGAGTGGCTGAAGATGAAGCCTGTTGGATT 480
 QY 585 AsnArgIleLysSerTrpAla 591
 DB 481 AGTCGATAGAGAAATGGGCA 501

RESULT 6

AAA89377/c
 ID AAA89377 standard; DNA; 519 BP.

XX AC AAA89377;

XX 11-SEP-2003 (revised)

DT 23-APR-2001 (first entry)

XX XX

DE Brassica oleracea silencing gene homologous DNA seq1-23.
 XX Gene silencing; silencing gene; MOM; ds.

XX Brassica oleracea; acephala.

XX WO200100801-A2.

XX 04-JAN-2001.

PD

XX 21-JUN-2000; 2000MO-EP005761.
 PF 23-JUN-1999; 99GB-00014623.
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 PI WPI; 2001-137952/14.
 DR Novel gene encoding a protein that controls gene silencing, in particular
 PT silencing of plant genes.
 PT Example 6; Page 44; 48pp; English.
 PS This is the sequence of Seq1-23, a genomic DNA sequence of Brassica
 CC oleracea acephala that shows homology to the newly identified silencer
 CC gene (see AAA89353) of Arabidopsis thaliana. Seq1-23 is a partial
 CC sequence of clone 1, which was isolated from a genomic library screened
 CC with MOM cDNA. Partial sequences of clone 1 (see also AAA89378-80) showed
 CC similarity to different regions of the MOM gene (80-86% at DNA level and
 CC 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-
 CC terminal parts of the MOM protein (see BAB20062). Gene silencing is
 CC useful as a molecular tool for regulating gene expression. (Updated on 11
 CC -SEP-2003 to standardise OS field)
 XX SQ Sequence 519 BP; 150 A; 117 C; 101 G; 149 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 6.25e-33 Length: 519
 Score: 442.00 Matches: 89
 Percent Similarity: 84.82% Conservativity: 6
 Best Local Similarity: 79.46% Mismatches: 13
 Query Match: 13.74% Indels: 4
 DB: 4 Gaps: 1

US-10-018-929C-3_COPY_650_1270 (1-621) x AAA89377 (1-519)

QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
 DB 324 CGCGCTGAGTCTTTTATACCTTTTGGAGAGCCTGAATCCATCCATGACGTTAAGTTC 265
 QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
 DB 264 TTAGAGAAGATAAAGGTTGAGTGTCTTCTCAAGAACTAAGATTTTCCGGTTGTACTCA 205
 QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
 DB 204 GTGTTTACAGTTGAAGAAAAAACCCCTGATTTCTTCT-----AGGCAAAATAAG 157
 QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGluValAspTyr 80
 DB 156 CCTCTAGAGAACCTAAACCGCTCTCACATGATTCATGCTGGGGGCGCATCATAT 97
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 96 TTATTTGATAAGCTGGAACACTTCCATGGGAGTGGAAACCCAGGTTTCAGGAGTTCCATT 37
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
 DB 36 GAACAGTCTATTAAAGGATGCGGTAANNACAGGAATTC 1

RESULT 7

ADQ59166

ID ADQ59166 standard; cDNA; 5220 BP.

XX AC ADQ59166;

XX 09-SEP-2004 (first entry)

XX MSI-H carcinoma cDNA sequence SEQ ID NO:3.

DE

XX coding mononucleotide repeat; cMNR; gene; antibody; MSI-H tumour;
 KW MSI-H carcinoma; high microsatellite instability tumour;
 KW high microsatellite instability carcinoma; cytostatic; ss.
 XX Homo sapiens.
 XX KR2004008012-A.
 XX 28-JAN-2004.
 XX 15-JUL-2002; 2002KR-00041304.
 XX 15-JUL-2002; 2002KR-00041304.
 XX (KIMH/) KIM H G.
 XX (KIMN/) KIM N G.
 XX (LEEJ/) LEE J S.
 XX (RHEE/) RHEE H S.
 XX Kim HG, Kim NG, Lee JS, Rhee HS;
 XX WPI; 2004-386326/36.
 XX Genes containing coding mononucleotide repeats are useful in developing
 PT an antibody against MSI-H (high (sic high) microsatellite instability)
 PT tumor.
 XX Claim 1; SEQ ID NO 3; 578pp; Korean.
 XX The present invention describes genes containing coding mononucleotide
 CC repeats (cMNRs). The genes are useful for the development of an antibody
 CC against MSI-H (high microsatellite instability) tumour. Also described:
 CC (1) cDNA genes containing cMNRs with 10 or more nucleotide sequences, and
 CC selected from the cDNA genes having the nucleotide sequences of SEQ ID
 CC NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35,
 CC 37, 39, 41 and 43; (2) cDNA genes, which are frameshift mutated by
 CC deletion or insertion of one or more base in the cMNRs; (3) genomic DNA
 CC genes containing cMNRs with 10 or more nucleotide sequences, and selected
 CC from the genomic DNA genes having the nucleotide sequences of SEQ ID
 CC NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36,
 CC 38, 40, 42 and 44; (4) genomic DNA genes, which are frameshift
 CC mutated by deletion or insertion of one or more base in the cMNRs. The
 CC genes have cytostatic activity. The present sequence represents an MSI-H
 CC carcinoma cDNA sequence from the present invention.
 XX
 SQ Sequence 5220 BP; 1766 A; 1025 C; 1366 G; 1063 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 6.58e-05 Length: 5220
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conservative: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: 12 Gaps: 37

US-10-018-929C-3_COPY_650_1270 (1-621) x AD059166 (1-5220)

Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
 Db 2635 GCGGACACAGTCGTCATCTTGACTCTGACTGGAAACCCACAGATGACTTGCAGGCACAA 2694
 Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 Db 2695 GCCCGCGCATAGAAATGGTCAGAGAAGCAGGTAATAATTTACCCTAGTTACAAAG 2754
 Qy 42 CysThrValGluGluLysAla-----Leu 49
 Db 2755 GGGACTGTGGAGGAGGAGATCATAGACGGGCCAAAAGAGATGGTATTAGATCATCTG 2814
 Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
 Db 2815 GTGATTGAGCGCATGGACACCACCTGCGCGGACGATCCTGGAAACAACACTCAGGAAGGTCC 2874

Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 2875 AACTCAAAATCCTTTTAAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGAT 2934
 Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 2935 CTCTTCAAGAACTGGNA-----GGGAGGAATCA 2964
 Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
 Db 2965 GAACCTCAGGAAATGGAT-----ATAGATGAAATTTTGGCGTTGGCTGAA 3009
 Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuAlaLysHisAlaGlnGlyThr 140
 Db 3010 ACGAGAGAGAAATGAAGTGTCAACAGTGCAACAGATGAATCTTCTA---TCACAGTTTAA 3066
 Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 Db 3067 GTTGCCCACTTTCGAACATGGGAAGATGAAGAGAGCTA-----GAAGAGCGT 3114
 Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 Db 3115 CCTCACAAGGACTGGGATGAGATCATT-----3141
 Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 Db 3142 -----CCAGAGGAACAAGGAAAGTAGAGGAGGAGCGGCGCAAGAGGAGCTAGAA 3195
 Qy 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3196 GAAATTTATATGCTGCTCGAATTCGAGGTTCCTCACTAAAAGGCTCAGACAAATGACAGT 3255
 Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
 Db 3256 GACTCTGACACTGAGTCTAAGAGCGCAGCCACAGATCTCTCTCTCTGAGAGTGAACG 3315
 Qy 224 ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 Db 3316 GAAGAC-----TCTGATGATGACAAAG-3339
 Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 Db 3340 -----CCAAAGCGCAGAGCGCTCCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 3390
 Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 Db 3391 TTTACTGATGCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT-----3438
 Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 Db 3439 -----GGTCTCCCTCTTGACCGCTGGAGTGTGCTTACCGCTGAT---GCTGAGCTG 3486
 Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 3487 GTAGATAAGTCGCTGGCAGATCTG-----AAGCGCTCGGT 3522
 Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 Db 3523 GAACCTGATCCAC-----AACAGCTGTGTCTCAGCAATGCAAGATATGAAGAGCAGCTG 3576
 Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
 Db 3577 AAAGAAAATCCCGAGGAGGAGGAAAGACCGGAGGAAAAGG---AGAGTCCACATATCAAG 3633
 Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 3634 ATATCCGAGCTT---CAGGTTAATGTGAATCCATTCCATCAACATGAAGAGGAGTTT---3687
 Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 3688 -----GAGATGCTGCATAAATCTATCCCTGTGAGCCCTCGAAGAAAAA 3735

Qy	395	TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe	414
Db	3736	TACTGCTTAACCTCTCGTGTCAAAAGCTGCACATTTTTCATGTAGAGTGGGGGTGAAGAT	3795
Qy	415	AspCys-----PheGly-ThrAsnSer-----Ly	422
Db	3796	GATTCTCGCTTGTCTGGGGATTATTAAACATCGCTATGAAACTGGGAGTTAATTAAA	3855
Qy	422	sGlnSerValValSerThrLysLeuValAsn-----	432
Db	3856	ACAG-----ACCCAGAGCTTAAATTAACCTGACAAATTCCTCGCGTGGAGACA	3903
Qy	433	-----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet	450
Db	3904	ATAAAAGCCTCAGGGGAAGCAGCTACAGACCCGAGCGGATTACTTGTGTGAAGAGTCTC	3963
Qy	451	ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys	463
Db	3964	AGAAAGGCTGGAGAAAGAGGGGGCTGTGCACAGTGGGAGGAGGCCAAATTAAGNAG	4023
Qy	464	Arg-----	464
Db	4024	CGGAAGCCTCGGTTAAAGAAAGAAAAACAAAGTGCCAGGCTGAAAGAGGAGCATGGAATT	4083
Qy	465	-----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
Db	4084	GAGCTTTCATCTCTAGGCAATTCAGATAATCCATCAGAGAGGAGAGTGAAGATGAT	4143
Qy	481	GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
Db	4144	GGCTTGGAAAAAAGTCCAAATGAAAAAACAAGAG---AAGAAAGAGAACAGGAGAAC	4200
Qy	501	LysMetGluLeuLeuAsnMetTyrAlaAspLys-----	511
Db	4201	AAGAGAGAAACAATAGATTCTTAGAAAGACACAAGAGGGGACAGGAAGAAAGAGTCA	4260
Qy	512	LysGlnLysLeuGluThrSerLysSerValGluAla-----	523
Db	4261	AAAGATAAGAAAGAGAGAGCCGTAAAGGTGTGATGCCAAATCTTCGAGTAAATCAAAGCGA	4320
Qy	524	-----AlaValIleAArgIleThrCysSerArgThrSerThrGlnValGlyAsp----	539
Db	4321	TCTCAGGGTCTGTCCATATTACACAGAGGAAGTGAACCTGTCCCATTTGGAGAGGATGAG	4380
Qy	539	-----	539
Db	4381	GATGATGATCTGGACACGAGACACATTTCAGCATATGTAAGGAGAGGATGAGGCCCGTGAAA	4440
Qy	540	-----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp	552
Db	4441	AAGGCACTGAAACAGCTCGACAAACCTGACAGAGGGGCTCAACGTGCAAGAACAGCTGGAA	4500
Qy	553	-----GluIleLysSerGluLysAsnGluCysLeuLysSer----	564
Db	4501	CACACCGGAACCTGCTGTGAAATTCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC	4560
Qy	564	-----	564
Db	4561	TCAGATCAGGACACATCAAACTCTGAGGAGGAGAACCTATGGATTTTGTTCNAGTTT	4620
Qy	565	-----LeuGlnGlnMetHisGluValAlaLysLysLeuAla	577
Db	4621	ACAGATTGTATGCTCGAAAAACTGCATAAAGTTATACAAGATGGCTCATAGAAGAGGTCT	4680
Qy	578	GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys	597
Db	4681	CAAGAAGAAGAG-----GAGCAAAAAGAAAGAAAGACGAC	4713
Qy	598	ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn	613
Db	4714	GTG-----ACTGGGGTAAAGAAACCATTTTCGTTCAGAGGGCTCAGGCTCCAGCGG	4764
Qy	614	-----IleSerGlnAsn	617

Db 2760 GCGGACACAGCTCGTCATCTTTGACTCTGACTGGAAACCCAGCAATGACTTTCAGGCACAA 2819
 Qy 22 GluValIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 Db 2820 GCCCGAGCGCATAGAAATGGTCTCAGAAAGACAGGTAATATTATTCACCGCTTAGTTACAAAG 2879
 Qy 42 CysThrValGluGluLysAla-----Leu 49
 Db 2880 GGGACTGTGGAGGAGGAGATCATAGAACGGGCCCAAAAGAGAGATGGTATTAGATCATCTG 2939
 Qy 50 IleLeuAlaArgGlnAenMetArgGlnAsnLysAlaValGluAen----- 64
 Db 2940 GTGATTTCAGCGCATGGACACCACTGGCGGACGATCCTGGAAACAACTCAGGAAGGTCC 2999
 Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 3000 AACTCAAAATCTTTAATAAAGAGAGCTGACAGCTATTTTGAAATTTGGAGCAGAGGAT 3059
 Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 3060 CTCTTCAAGAACTGGAA-----GGGGAGGAATCA 3089
 Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
 Db 3090 GAACCTCAGGAATGGAT-----ATAGATGAAATTTTTCGGTTGGCTGAA 3134
 Qy 121 GlyGluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 3135 ACAGAGAGAGAATGAAGTGTCAACAAGTGCAACAGATGAACTTCTA---TCACAGTTTAAAG 3191
 Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 Db 3192 GTTGCCAACTTTGCAACAAATGGAGATGAAGAGAGCTA-----GAAGAGCGT 3239
 Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAenProMetTrpLysTyrProSer 180
 Db 3240 CCTCACAAGGACTGGGATGAGATCAAT----- 3266
 Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu----- 194
 Db 3267 -----CCAGAGAACAAAGAAAGAAAGTAGAGAGGAGGCGGACAGAGGAGCTAGAA 3320
 Qy 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3321 GAAATTTATATGTGCTCGAATTCGGAGTTCCTAAAGGCTCAGACAAATGACAGT 3380
 Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
 Db 3381 GACTCTGACACTGAGTCTAAGAGCGAGGCCAGAGATCCTCTGCTCTGAGAGTGAAACG 3440
 Qy 224 ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 Db 3441 GAAGAC-----TCTGATGATGACAAGAAG----- 3464
 Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 Db 3465 -----CCAAAGCCACAGGGCGTCCAGGAGTGTGCGGAGGACCTCGTGGAGGA 3515
 Qy 264 ThrLeuAspGlyAen-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 Db 3516 TTTACTGATCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT----- 3563
 Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 Db 3564 -----GGTCTCCCTCTTGAACCGCTCGAGTGCATGACCGTGAT---GCTGAGCTG 3611
 Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 3612 GTAGATAAGTCGTGGCAGATCTG-----AAGCGCTGGGT 3647
 Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAenPheLeuGluTyrVal 336
 Db 3648 GAACATGATCCAC-----AACAGCTGTGTGTACAGCAATGCAGGAATACGAAGACAGCTG 3701

Qy 337 IleGluAenHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
 Db 3702 AAGAANAATGCCAGCAGGAAAGACGCGGAAAAAGG---AGAGTCCAAACATCAAG 3758
 Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 3759 ATATCCGGAGTT---CAGGTTAATGTGAAATCCATTATCCAAATGAAGAGGAGTTT--- 3812
 Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 3813 -----GAGATGTCTGATAAATCTATCCTGTGGACCTCGAAGAAAAAATAA 3860
 Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 Db 3861 TACTGCTTAACCTGTCGTGTCAAAGCTGCACATTTTGTAGTGGGGGGTGGAAAGAT 3920
 Qy 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
 Db 3921 GATTCTCGCTGTGTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTTAATTA 3980
 Qy 422 sGlnSerValValSerThrLysLeuValAen----- 432
 Db 3981 ACAG-----ACCAGAGCTTAAATTAACCTGACAAAAATTCGCGGTGGAGACA 4028
 Qy 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
 Db 4029 GATAAAAGCCTCAGGGAGAGCAGCTACAGACCCGAGCGGATTTACTTGTGTGAAGCTGCTC 4088
 Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
 Db 4089 AGAAGGGTCTGGAGAAGAAGGGGGCTGTGACAGGTGGGGAAGCAGCCAAATTAAGAAG 4148
 Qy 464 Arg----- 464
 Db 4149 CGGAAGCCTCGGGTAAAGAAAGGAAAAACAAGTCCCAGGCTGAAAGAGGAGCATGGAATT 4208
 Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
 Db 4209 GAGCTTTTCATCTCTAGGCATTCAGATATATCCATCAGAGAGGAGAGAGTGAAGATGAT 4268
 Qy 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
 Db 4269 GGCTTCGAAAAAAGTCCAAATGAAAAAACAAGAAG---AAGAAAGAGAACAAAGGAGAAC 4325
 Qy 501 LysMetGluLeuLeuAenMetTyrAlaAspLys----- 511
 Db 4326 AAGGAGANAACAATGAGTTCTAGGAAAGACAAAGAAAGGGGACAAAGAAAGAAAGAGTCA 4385
 Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
 Db 4386 AAAGATAAGAAAGAGAGCCTAAAGTGGTGATGCCAATCTTCGAGTAATCAAGCGA 4445
 Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
 Db 4446 TCTCAGSGTCTCTCCATATTACAGCAGGAAGTGAACCTGTCTCCCATTTGGAGAGGATGAG 4505
 Qy 539 ----- 539
 Db 4506 GATGATGATCTGGACCAGGAGACATTCAGCATATGTAAGAGAGGATGAGCCCGTGAAA 4565
 Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
 Db 4566 AAGGCACTGAACAGCTCGCAACAACTGACAGGGGCTCAACGTGCAGAGACAGCTGGAA 4625
 Qy 553 -----GluIleLysSerGluLysAenGluCysLeuLysSer----- 564
 Db 4626 CACACCCGGAACCTGCTGCTGTAATAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCTAC 4685
 Qy 564 ----- 564
 Db 4686 TCAGATCAGGAGCACATCAAACTCTCGAGGAGGAACCTATGAGATTTTGTTCACAGTTT 4745

QY 565 -----LeuGluGlnMethHisGluValAlaLysLysLysLysLeuAla 577
Db 4746 ACAGAAATTGATGCTCGAAACTCATAGTTATACAGATGGCTCATAGAAAGGTCT 4805
QY 578 GluAspGluAlaCysTTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 4806 CAAGAAGAGAG-----GAGCAAAAGAGAAAGAGACAC 4838
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SergLysSerSerAsn 613
Db 4839 GTG-----ACTGGGGTGAAGAAACCATTTGCTCCAGAGGGCTCAGGCTCCAGCGG 4889
QY 614 -----IleSerGlnAsn 617
Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 9
ADL63336
ID ADL63336 standard; DNA; 7764 BP.
XX
AC ADL63336;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #21548.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN W0200170979-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
PT WPI; 2001-611502/70.
XX
PS Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 21548; 106pp; English.

CC The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein

CC fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. CC This sequence represents a human ovarian cancer DNA marker of the CC invention.

XX
SQ Sequence 7764 BP; 2342 A; 1534 C; 1892 G; 1979 T; 0 U; 17 Other;
Alignment Scores:
Pred. No.: 0.000118 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 5 Gaps: 37

US-10-018-929C-3_COPY_650_1270 (1-621) x ADL63336 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
Db 3342 GCGGACACAGTCGTCATCTTTGACTCTGCTGGAACCCCGAATGACTTGCAGGCACAA 3401
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3402 GCCCGAGCGCATAGAATTGGTCAGAAGACGACGTAATATTACCGCTTAGTTACAAG 3461
QY 42 CysThrValGluGluLysAla-----Leu 49
Db 3462 GGGACTGTGGAGGAGGAGATCATAGAACGGCCCAAGAGATGGTATTAGATCATCTG 3521
QY 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
Db 3522 GTGATTGAGCGCATGGACACCCTGGCCGACGATCTCTGGAACAACTCAGGAAGGTCC 3581
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3582 AACTCAAAATCCTTTTAATAAGAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGAT 3641
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3642 CTCTCAAGAACTGGAA-----GGGAGGAGATCA 3671
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3672 GAACCTCAGGAATGGAT-----ATAGATCAAAATTTTTCGGTTCGCTGAA 3716
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3717 ACAGAGAGAATGAAGTGTCAACAAGTGCAACAGATGAACCTCTA---TCACAGTTTAAAG 3773
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3774 GTTGCCCAATTTCCACAATGGAAGATGAAGAAGAGCTA-----GAAGACCGT 3821
QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 3822 CCTCACAAGGACTGGGATGAGATCAIT-----3848
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
Db 3849 -----CCAGAGGAACAAAGGAAAAAGTAGAGGAGGAGGAGGAGGAGGAGTAAAG 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3903 GAAATTTATATGCTGCTCGAAATTCGAGATTCCTCACTAAAGGCTCAGACAAATGACAGT 3962

```
QY 204 AspGlyGlyAsnAlaLysLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
DB 3963 GACTCTGACACTGAGCTTAGAGCGCAGCCCGACAGATCCTCTGCTCTGAGAGTGAACG 4022
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
DB 4023 GAAGAC-----TCTGATGATGACAAGAG----- 4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
DB 4047 -----CCAAAGCGCAGAGCGCGTCCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
DB 4098 TTACTCTGATCAGAGATCCGAAGGTTTCATCAAGGCTTTATAGAGTTT----- 4145
QY 279 HisLeuSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
DB 4146 -----GGTCTCCCTCTTGAACGCGTGGAGTGCATACGACGTGAT---GCTGAGCTG 4193
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
DB 4194 GTAGATAAGTCCGTGGCAGATCTG-----AAGCGCTGGGT 4229
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
DB 4230 GAATCTATCCAC-----AACAGTGTGTGTGCAATGAGCAATGAGCAAGAGCGCTG 4283
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
DB 4284 AAGAAAAATCCAGCGAGGAAAGGACCGAGGAAAGG---AGAGTCCCAACAATCAAG 4340
QY 357 LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
DB 4341 ATATCCGGAGTT---CAGGTAAATGTGAATCCATTATCCATTCACCATGAAGAGGATTT--- 4394
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
DB 4395 -----GAGATGTGTGCATAAATCTATCTCTGTGGACCTCGAAGAAAAA 4442
QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
DB 4443 TACTGCTTAACCTGCTGTGCTCAAAGCTGCACATTTTGTATGATGAGTGGGGGTGGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
DB 4503 GATTCTCGCCTGTTGCTGGGATTTTATGAACATGGCTATGGAAACTGGGAGTTAATTAA 4562
QY 422 sGlnSerValIleSerThrLysLeuValAsn----- 432
DB 4563 ACAG-----ACCCAGAGCTTAATAATTAACACAAAATTCGCGCGTGGAGACA 4610
QY 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
DB 4611 GATAAAAGGCTCAGGGAAGCAGCTACAGCCGAGCGGATTAATTTGTTGAGCTGCTC 4670
QY 451 ArgAsnSerSerGly-----AspGluGluCysMetThrGluLys 463
DB 4671 AGAAAGGGTCTGGAGAAGAAGGGGGCTGTGACAGGTGGGAAGAGGCCAAATTAAGAAG 4730
QY 464 Arg----- 464
DB 4731 CGGAAGCCTCGGTAAAGAGAGGAAAAAACAAGTGCCCGCAGGCTGAAGAGGAGCATGGAAAT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
DB 4791 GAGCTTTTCATCTCTAGGCATTTCAGATAATCCATCAGAGAGGGAGAGTGAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 4851 GGCTTGGAAAAAAGTCCATCAATCAAAAAAACAAGAG---AAGAAAGAGAACAAAGGAGAAC 4907
```

```
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
DB 4908 AAGGAGAAACAATAGATTCTAGGAAGACAAAGAGGGGCAAGGAAGAAAGAGTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
DB 4968 AAAGATAAGAAAGAGAGGCTTAAAGTGGTGATGCCAAATCTTTCGAGTAAATCAAAGCGA 5027
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
DB 5028 TCTCAGGGTCTGTCTCCATATTACAGCAGGAAGTGAACCTGTCTCCCATTTGGAGAGGATGAG 5087
QY 539 ----- 539
DB 5088 GATGATGATCTGGACACGAGGAGACATTCAGCATATGTAAAGAGAGGATGAGCCCGTGAAA 5147
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
DB 5148 AAGGCACATGAACAGCTCGACAAACCTGACAAAGGGCTCAACGTGCAAGAACACAGCTGGAA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
DB 5208 CACACCCGGAACCTGCTGTGTAATTCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
QY 564 ----- 564
DB 5268 TCAGATCAGGAGCACATCAAACTCTCGGAGGAGAACCTATGGATTGTTTCTTCCAAGTTT 5327
QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
DB 5328 ACAGAATTTGATGCTCGAAACTGCAATAAGTTATACAGATGCTCATAGAAGAGGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
DB 5388 CAAGAAGAGAG-----GAGCAAAAGAGAAAGAGACGAC 5420
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
DB 5421 GTG-----ACTGGGGGTAAAGAACCATTTCTGTCAGAGAGCCTCAGGCTCCAGCCGG 5471
QY 614 -----IleSerGlnAsn 617
DB 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 10
AAD54628
ID AAD54628 standard; DNA; 7764 BP.
XX
AC AAD54628;
XX
DT 26-JUN-2003 (first entry)
XX
DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #3.
XX
KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytosstatic; ds.
XX
OS Homo sapiens.
XX
PN WO200298899-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017466.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX
```



```
Qy 464 Arg----- 464
Db 4731 CGAAGCCTCGGTAAAGAGGAAACAAAGTGCCAGGCTGAAAGAGGAGCATGGAAATT 4790
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4791 GAGCTTTCATCTCCTAGGCATTCAGATAATCCATCAGAGGAGGAGGAGTGAAGATGAT 4850
Qy 481 GlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4851 GGCTTGGAAAAAGTCCAATGATAAAAAAACAAGAGAA--AAGAAAGAGAAACAAGGAGAAC 4907
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4908 AAGGAGAAACAATAGTCTTAGGAAAGACAAAGAGGGGACCAAGGAAAGAAAGAAAGTCA 4967
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4968 AAGATGAAGAAAGAGAGCCTTAAGTGGTGATGCCAAATCTTCGAGTAATCAAGCGA 5027
Qy 524 -----AlaValIleArgIleTyrCysSerArgThrThrGlnValGlyAsp----- 539
Db 5028 TCTCAGGGTCTGTCCTCATATTACAGCAGGAAGTGAACCTGTGCCCATTCGAGAGGATGAG 5087
Qy 539 ----- 539
Db 5088 GATGATGATCGGACCGAGAGACATTCAGCATATGTAAAGAGAGGATGAGGCCCGTGAAA 5147
Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5148 AAGCACTGAACAGCTCGACAAACCTGACAAAGGGGCTCAACGTGCAAGACAGCTGGAA 5207
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5208 CACACCGGAACCTGCTGTGTAATCGGAGCCGATAGCCGAGTGCCTTAAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGGAGCACATCAAACTCGGAGGAGAACCTATGATGATTTTTCACAGTTT 5327
Qy 565 -----LeuGluGlnMethHisGluValAlaLysLysLysLeuAla 577
Db 5328 ACAGATTGTGCTCGAAACCTGCATAGTATTACAGATGGCTCATAGAAAGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 5388 CAAGAAGAAGAG-----GAGCAAAAGAGAAAGAGAGCAG 5420
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGGTGAAGAACCAATTCGTCAGAGGCGCTCAGGCTCCAGCGG 5471
Qy 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492
```

RESULT 11

```
ADK60458
ID ADK60458 standard; DNA; 7764 BP.
XX
AC ADK60458;
XX
XX 06-MAY-2004 (first entry)
XX
DE Angiogenesis differentially expressed gene GS-N32.
XX
KW ds; vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
KW antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;
KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovary hyperstimulation; prolasia; endometriosis; restenosis;
KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
```

```
KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
KW ischemia; angina; myocardial infarction; chronic heart disease;
KW cardiac congestion; macular degeneration; osteoporosis.
OS Homo sapiens.
XX
PN FR2836687-A1.
XX
PD 05-SEP-2003.
XX
PF 11-APR-2002; 2002FR-00004546.
XX
PR 04-MAR-2002; 2002FR-00002717.
XX
XX (GENE-) GENE SIGNAL.
XX (ALMA/) AL MAHMOOD S.
XX
XX Colin S, Schneider C, Al Mahmood S;
XX
XX WPI: 2004-013912/02.
XX P-PSDB; ADK60208.
XX
XX Compositions for diagnosing, prognosing and treating angiogenic disorders
XX including tumor vascularization and heart disease, comprise nucleic acid
XX or polypeptide differentially expressed in angiogenesis.
XX
XX Claim 2; SEQ ID NO 33; 424pp; French.
XX
XX The invention relates to a novel pharmaceutical composition active on
XX angiogenesis comprising an endothelial cell nucleic acid whose expression
XX is induced by an angiogenic factor and inhibited by an angiostatic agent
XX or its complement or fragment, a polypeptide sequence encoded by the
XX nucleic acid or its fragment, a molecule capable of inhibiting expression
XX of the nucleic acid or a molecule which binds to the polypeptide
XX sequence. The invention is used to diagnose, prognose or treat an
XX angiogenic disorder in a mammal, particularly a human. The disorder is
XX particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
XX Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
XX endometriosis associated with neovascularization, restenosis due to
XX angioplasty, overproduction of tissue due to cicatrization, a peripheral
XX vascular disease, hypertension, vascular inflammation, Raynaud disease,
XX aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
XX myocardial infarction, chronic heart disease, cardiac congestion or
XX macular degeneration due to age or osteoporosis. This sequence
XX corresponds to a differentially expressed DNA used in the composition of
XX the invention.
XX
SQ Sequence 7764 BP; 2342 A; 1535 C; 1898 G; 1989 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 0.000118 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 37 Gaps: 37
```

US-10-018-929C-3_COPY_650_1270 (1-621) x ADK60458 (1-7764)

```
Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
Db 3342 GCGGACACAGCTCGTCATCTTTGACTCTGCTGGAACCCCGAGAAATGACTTGCAGGCACAA 3401
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3402 GCCCGAGCGCATAGAATTGGTCAGAGAGACAGGTAATATTTACCGCTTAGTTTACCAAG 3461
Qy 42 CysThrValGluGluLysAla-----Leu 49
Db 3462 GGGACTCTGGAGGAGGAGATCATAGAACGGGCCCAAAAGAGATGGTATTAGATCATCTG 3521
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn----- 64
```

Db 3522 GTGATTCAGCCATGGACACCACTGGCGGACGATCTCTGGAAACAACACTCAGGAGGTCC 3581
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3582 AACTCAAAATCTTTAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGAT 3641
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3642 CTCTTCAAGAACTGGAA-----GGGAGGAAATCA 3671
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTGGGTTGGCTGAA 3716
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3717 ACGAGAGAGATGAAGTGTCAACAAGTCAACAGACTTCTA-----TCACAGTTTAAG 3773
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3774 GTTGCCAACTTTCGCAAAATGGAAAGATGAAAGAGCTA-----GAAGAGCGT 3821
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 3822 CTTCAAGAGACTGGATGAGATCTT----- 3848
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTrpPheGlu----- 194
Db 3849 -----CCAGGAGAACAAAGGNAAGATAGAGGAGGAGCGGAGAGGAGCTAGAA 3902
Qy 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3903 GAAATTTATATGTCCTCGAATTCGAGTTTCCACTAAAGGCTCAGACAAATGACAGT 3962
Qy 204 AspGlyGlyAsnAlaLysLysArgLysAlaSerAspValThrAspProArgVal 223
Db 3963 GACTCTGACACTGAGTCTAGAGCAGCCAGAGATCCTCTGCTCTGAGAGTGAACG 4022
Qy 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 4023 GAAGAC-----TCTGATGATGACAAGAG----- 4046
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 4047 -----CCAAAGCGCAGAGGCGCTCGAGGAGTGTGCGGAGGACCTCGTGGAGGA 4097
Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 4098 TTTACTGATCAGAGATCCGAGGTTTCATCAAGGCTTATGAAGATTT----- 4145
Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db 4146 -----GGTCTCTCTTGAACGCTGGAGTGTGCTTAGCACCTGAT---GCTGAGCTG 4193
Qy 297 ProAspGluSerGlnArgLysLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4194 GTAGATAAGTCCGTGGCAGATCTG-----AAGCGCTCGGT 4229
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 4230 GAACTGATCCAC-----AACAGTGTGTGTCAGCAATGAGGAATATGAAGAGCAGCTG 4283
Qy 337 IleGluAsnHisArgIleTyrGluLupProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 4284 AAAGAAATGCCAGGAGGAAAGGACCGGNAAGG---AGAGGTCCAAACATCAAG 4340
Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 4341 ATATCCGAGTT---CAGGTTAATGTGAATCCATTATCCAAACATGAAGAGGAGTT--- 4394
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db 4395 -----GAGATGCTGCATAAATCTATCTCTCTGAGCCCTCGGAGGAAAAA 4442

Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 4443 TACTGCTTAACCTGTGTCGTCAAAAGCTGCACATTTTGTATAGAGTGGGGTGGAAAGAT 4502
Qy 415 AspCys-----PheGly-ThrAsnSer---Ly 422
Db 4503 GATTCCTCCCTGTTGCTGGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATAAA 4562
Qy 422 sGlnSerValIleSerThrLysLeuValAsn----- 432
Db 4563 ACAG-----ACCCAGAGCTTAATAATTAACAGCAAAATTCCTCCCGTGGAGACA 4610
Qy 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db 4611 GATAAAAGCTCAGGGAAGACAGTACAGACCCGAGCGGATTACTTTGTTGAAGCTGCTC 4670
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4671 AGAAGGCTCTGGAGAAGAGGGGGCTGTGCACAGTGGGAGGAGGCCAAATTAAGAAG 4730
Qy 464 Arg----- 464
Db 4731 CGGAGCCTCGGCTAAAGAGGAAACAAAGTCCAGGCTGAAAGAGGAGCATGGAATT 4790
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4791 GAGCTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGAGGAGTGAAGATCAT 4850
Qy 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4851 GGCTTGGAAAAAGTCCAAATGAAAAAAGAG---AAGAAAGAGAAACAAGGAGAAC 4907
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4908 AAGGAGAAACAATGAGTTCTAGGAAAGACAAAGAGGGGACAAAGAAAGAAAGAACTCA 4967
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4968 AAAGATAAGAAAGAGAGCCCTAAAGTGGTGTGATGCCAAATCTTCGAGTAATCAAAGCGA 5027
Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp- 539
Db 5028 TCTCAGGCTCTGTCCATATTACAGCAGGAAGTGAACCTCTCCCATTTGGAGAGGATGAG 5087
Qy 539 ----- 539
Db 5088 GATGATGATCTGGACCCAGGAGACATTCAGCATATGTAAAGGAGAGGATGAGCCCGTGAA 5147
Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5148 AAGGCACTGAACAGCTCGACAAACCTGACAAAGGGGTCAACGTCGAAGAACAGCTGGAA 5207
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5208 CACACCCGGAACCTGCCTGTGTAATAATGGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGGAACTATGATGATTTTGTTCCAAGTTT 5327
Qy 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 5328 ACAGAAATTTGATGCTCGAAAACTGCATAGTATACAAGATGGCTCATAGAAGAGGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
Db 5388 CAAGAGAGAGAG-----GAGCAAAAGAAAGAAAGACGAC 5420
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGTAAAGAAACCAATTTTCGTCCAGAGGCTCAGGCTCCAGCGG 5471

QY	614	-----IleSerGlnAsn 617	Best Local Similarity: 20.02%	Mismatches: 251
DB	5472	GACTCTGTATATCTCAGTCC 5492	Query Match: 5.16%	Indels: 283
			DB: 12	Gaps: 37
RESULT 12				
ADK60759				
ID	ADK60759	standard; DNA; 7764 BP.		
AC	ADK60759;			
XX				
DT	06-MAY-2004	(first entry)		
DE	Angiogenesis differentially expressed gene GS-N32.			
XX				
KW	ds; vasotropic; antirheumatic; antiarthritic; hypotensive; antianigmal;			
KW	antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;			
KW	angiogenesis; endothelial cell; diagnosis; tumor vascularization;			
KW	retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;			
KW	ovary hyperstimulation; psoriasis; endometriosis; restenosis;			
KW	angioplasty; cicatrization; peripheral vascular disease; hypertension;			
KW	vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;			
KW	ischemia; angina; myocardial infarction; chronic heart disease;			
KW	cardiac congestion; macular degeneration; osteoporosis.			
XX				
OS	Homo sapiens.			
XX				
PN	FR2836686-A1.			
XX				
PD	05-SEP-2003.			
XX				
XX	04-MAR-2002; 2002FR-00002717.			
PF				
XX				
PR	04-MAR-2002; 2002FR-00002717.			
XX				
PA	(GENE-) GENE SIGNAL.			
PA	(ALMA/) AL MAHMOOD S.			
XX				
PI	Colin S, Schneider C, Al Mahmood S;			
XX				
DR	WPI; 2004-013911/02.			
DR	P-PSDB; ADK60509.			
XX				
PT	Compositions containing nucleic acid or polypeptide differentially			
PT	expressed in angiogenesis are useful to diagnose, prognose and treat			
PT	angiogenic disorders including tumor vascularization and heart disease.			
XX				
PS	Claim 2; SEQ ID NO 33; 405pp; French.			
XX				
CC	The invention relates to a novel pharmaceutical composition active on			
CC	angiogenesis comprising an endothelial cell nucleic acid whose expression			
CC	is induced by an angiogenic factor and inhibited by an angiostatic agent			
CC	or its complement or fragment, a polypeptide sequence encoded by the			
CC	nucleic acid or its fragment, a molecule capable of inhibiting expression			
CC	of the nucleic acid or a molecule which binds to the polypeptide			
CC	sequence. The invention is used to diagnose, prognose or treat an			
CC	angiogenic disorder in a mammal, particularly a human. The disorder is			
CC	particularly tumor vascularization, a retinopathy, rheumatoid arthritis,			
CC	Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,			
CC	endometriosis associated with neovascularization, restenosis due to			
CC	angioplasty, overproduction of tissue due to cicatrization, a peripheral			
CC	vascular disease, hypertension, vascular inflammation, Raynaud disease,			
CC	aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,			
CC	myocardial infarction, chronic heart disease, cardiac congestion or			
CC	macular degeneration due to age or osteoporosis. This sequence			
CC	corresponds to a differentially expressed DNA used in the composition of			
CC	the invention.			
XX				
SQ	Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;			
XX				
Alignment Scores:				
Pred. No.:	0.000118	Length:	7764	
Score:	166.00	Matches:	162	
Percent Similarity:	33.99%	Conservative:	113	

US-10-018-929C-3_COPY_650_1270 (1-621) x ADK60759 (1-7764)				
QY	2	AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21		
DB	3342	GGGACACACAGTCGTCATCTTTGACTCTGACTGGAACCCCGAGTGTGAGGACAA 3401		
QY	22	GlulysIleLysIleGluSerCysSerGluArThrLysIlePheArgLeuTyrSerVal 41		
DB	3402	GCCCGAGCGCATAGAATTGGTCAGAGAAGCAGGTAATAATTTACCGCTTAGTTACAAG 3461		
QY	42	CysThrValGluGluLysAla-----Leu 49		
DB	3462	GGGACTGTGGAGGAGGAGATCATAGAACGGGCCCAAGAGATGGTATTAGATCATCTG 3521		
QY	50	IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64		
DB	3522	GTGATTCAGCGCATGACACACCCTGCCGGACCATCTCTGGAAAAACAACCTCAGGAGGTCC 3581		
QY	65	-----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80		
DB	3582	AACTCAAAATCCTTTTAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGGAT 3641		
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100		
DB	3642	CTCTTCAAGAAGACTGGAA-----GGGAGAGGAATCA 3671		
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120		
DB	3672	GAACCTCAGGAATGGAT-----ATAGATCAAAATTTGCGTTGGCTGAA 3716		
QY	121	GlyGluGluAsnGluValLysLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140		
DB	3717	ACGAGAGAGAAATGAAGTGTCAACAGTGCAACAGATGAACCTCTA---TCACAGTTTAAG 3773		
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer 160		
DB	3774	GTTGCCAACTTTGCCAACAATGGGAAGATGAAGAAGACTA-----GAAGAGCGT 3821		
QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180		
DB	3822	CCTCACAAGCACTGGGATGAGATCATT-----3848		
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194		
DB	3849	-----CCAGAGGACAAAGGAAAAAGTAGAGGAGGAGCGCGCAGAGGAGCTAGAA 3902		
QY	195	-----GlySerGluAlaSerProLysThrGly 203		
DB	3903	GAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAAGAGCTCAGCAAAATGACAGT 3962		
QY	204	AspGlyGlyAsnAlaLysArgLysAlaSerAspAspValThrAspProArgVal 223		
DB	3963	GACTCTGACACTGAGCTTAAGAGGAGGCCAGGATCTCTCTCTCTGAGAGTGAACG 4022		
QY	224	ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243		
DB	4023	GAAGAC-----TCTGATGATGACAAGAG-----4046		
QY	244	LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263		
DB	4047	-----CCAAAGCGCAGAGGGCTCGAGAGAGTGTGCGGAAGGAGCACTCTGCGGAGGA 4097		
QY	264	ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278		
DB	4098	TTTACTGATGCAGAGATCCGAAGTTTCATCAAGGCTTATAAGAAGTTT-----4145		
QY	279	HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296		
DB	4146	-----GGTCTCCCTCTTGAACCGCTGAGTGTCTTAGCAGCTGAT---GCTGAGCTG 4193		

Qy	297	ProAspGluSerGlnA-ArgLeuHisThrValLeuLysProLysMetalalysLeuCys	316
Db	4194	GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCCTGGGT	4229
Qy	317	GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal	336
Db	4230	GAATCGATCCAC-----AACCGTGTGTGTGCACATATGAGGAAGACAGCTG	4283
Qy	337	IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla	356
Db	4284	AAAGAAATGCCACGCGAGGNAAGGACCGGGNAAGG---AGAGGTCCACACATCAAG	4340
Qy	357	LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal	376
Db	4341	ATATCCGGAGTT---CAGGTTAATGTGAAATCCATTATCCAAATGAAGAGGAGTTT---	4394
Qy	377	ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle	394
Db	4395	-----GAGATGCTGCATAAATCTATCCCTGTGGACCTGAAGAAAAA	4442
Qy	395	TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe	414
Db	4443	TACTGCTTAACCTGCTGTGTCAAAGCTGCACATTTGATGTAGAGTGGGGGTGGAGAT	4502
Qy	415	AspCys-----PheGly-ThrAsnSer-----Ly	422
Db	4503	GATTCTCGCCTGTTGCTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATTAA	4562
Qy	422	sGlnSerValValSerThrLysLeuValAsn-----	432
Db	4563	ACAG-----ACCCAGAGCTTTAAATTAACACAAAATTCGCGGTGGAGACA	4610
Qy	433	---GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet	450
Db	4611	GATAAAGCCTCAGGGGAAGCAGCTACAGCCGACCGGATTACTTGTGAAGCTGCTC	4670
Qy	451	ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys	463
Db	4671	AGAAAGGCTCTGGAGAGAAAGGGGGCTGTGCAGGTGGGAGGAGGCCAAATTAAGAAG	4730
Qy	464	Arg-----	464
Db	4731	CGGAAGCCTCGGTAAGAAAGAAACAAAGTGCACAGGTGAAAGAGGAGCATGGAATT	4790
Qy	465	-----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
Db	4791	GAGCTTTCATCTCCTAGCGATTCAGATAATCCATCAGAGAGGGAGAGTGAAGATGAT	4850
Qy	481	GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
Db	4851	GGCTTGGAAAAAGTCCCAATGAAAAAACAAGAG---AAGAAGAGAACACAGGAGAAC	4907
Qy	501	LysMetGluLeuLeuAsnMetTyrAlaAspLys-----	511
Db	4908	AAGGAGAAACAAATGAGTTCTAGGAAAGACAAAGAGGGGACAAAGAAAGAAAGTCA	4967
Qy	512	LysGlnLysLeuGluThrSerLysSerValGluAla-----	523
Db	4968	AAAGATAAGAAAGAGAACGCTTAAAGTGGTGATGCCAAATCTTCGAGTAAATCAAGCGA	5027
Qy	524	-----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----	539
Db	5028	TCTCAGGGTCTGTCCATATATACAGCAGGAAGTGAACCTGTCCCATTTGGAGGATGAG	5087
Qy	539	-----	539
Db	5088	GATGATGATCTGGACAGGAGACATTCAGCATATGTAAAGGAGGATGAGGCCCGTGAAA	5147
Qy	540	-----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp	552
Db	5148	AAGGCACCTGAACAGCTCGACAAACCTTGACAGAGGGGCTCAACGCTGCAGAACACCTGGAA	5207

Qy	553	-----GlulleLysSerGluLysAenGluCysleuLysSer---	564
ADP73081		ADP73081	
Db	5208	CACACCGGAAGTGCCTGCTGAAATCGGAGACCGGATAGCCGAGTACCGCTTAAAGCCTAC	5267
Qy	564	-----	564
Db	5268	TCAGATCAGGAGCACATCAAACTCTCGAGGAGGAACCTATGGATTTTGTTCCTCAAGTTT	5327
Qy	565	-----LeuGluGlnMethIaGluValAlaLysLysLysLeuAla	577
Db	5328	ACAGAAATTTGATGCTCGAAACTGCATAAGTTATACAAGATGGCTCATAGAAGAGGTCT	5387
Qy	578	GluAepGluAlaCysTrpIleAenArgIleLysSerTrpAlaAlaLysLysLysValCys	597
Db	5388	CAAGAAGACAG-----GAGCAAAAGAAGAAGACGAC	5420
Qy	598	ValProIleGlnSerGlyAenAenLysHiePhe-----SerGlySerSerAen	613
Db	5421	GTG-----ACTGGGGGTAGAAACCATTTTCGTCAGAGCGCTCAGGCTCCAGCCGG	5471
Qy	614	-----IleSerGlnAen	617
Db	5472	GACTCTCTGATATCTCAGTCC	5492
RESULT 13			
ADP73081		ADP73081	
XX	AC	ADP73081; standard; DNA; 7764 BP.	
XX	DT	12-AUG-2004 (first entry)	
XX	DE	Angiogenesis inhibitor human DNA sequence, GS-N32.	
XX	KW	Inhibitor; angiogenesis; antisense nucleic acid; immunisation;	
XX	KW	angiogenic disorder; antiangiogenic; angiogenesis stimulator; cystostatic;	
XX	KW	dermatological; antiarthritic; antirheumatic; antiinflammatory;	
XX	KW	vasotropic; hypotensive; ophthalmological; antipsoriatic; cardiant;	
XX	KW	gene therapy; antisense gene therapy; tumour vascularisation;	
XX	KW	retinopathies; rheumatoid arthritis; Crohn's disease; atherosclerosis;	
XX	KW	ovarian hyperstimulation; psoriasis; endometriosis; restenosis;	
XX	KW	tissue granulation; peripheral vascular disorder; hypertension;	
XX	KW	vascular inflammation; Raynaud's disease; aneurism; arterial restenosis;	
XX	KW	thrombophlebitis; lymphadenopathy; lymphedema; ischaemia; angina;	
XX	KW	myocardial infarction; chronic heart disease; congestive heart disease;	
XX	KW	macular degeneration; human; ds; gene.	
OS		Homo sapiens.	
XX	FN	FR2843753-A1.	
XX	PD	27-FEB-2004.	
XX	PF	20-JUN-2003; 2003FR-00007507.	
XX	PR	04-MAR-2002; 2002FR-00002717.	
XX	PA	(GENE/) GENE S.	
XX	PA	(ALMS/) AL M S.	
XX	PI	Colin S, Schneider C, Al MS;	
XX	PI	WPI; 2004-216677/21.	
XX	DR	P-PSDB; ADP73132.	
XX	PT	Antisense nucleic molecule useful as inhibitor of angiogenesis in the	
XX	PT	treatment of angiogenic disorders, e.g., rheumatoid arthritis,	
XX	PT	atherosclerosis and endometriosis.	
XX	PS	Claim 1; SEQ ID NO 33; 404pp; French.	
XX	CC	The invention relates to a novel inhibitor of angiogenesis comprising an	
XX	CC	active substance chosen from at least one of a nucleic acid molecule, an	

CC antisense nucleic acid molecule, a polypeptide or an antibody. The
 CC invention further comprises: an antisense nucleic acid sequence chosen
 CC from any of the sequences provided in the specification; preparation of
 CC the antibody comprising *in vivo* or *in vitro* immunisation of an
 CC immunocompetent animal cell, preferably of a vertebrate and most
 CC preferably of a mammal, with at least one of the polypeptide sequences
 CC chosen from a sequence provided in the specification; a mammalian
 CC expression vector comprising at least one antisense sequence chosen from
 CC an antisense nucleic acid provided in the specification; preparation of a
 CC genetically modified cell, that over- or under-expresses a gene
 CC implicated in an angiogenic disorder, comprising inserting the vector
 CC from above into a mammalian cell; a genetically modified cell that over-
 CC expresses or under-expresses at least one gene involved in angiogenesis
 CC by a nucleic acid sequence chosen from any of ADP73049 to ADP73338, as in
 CC the antisense nucleic acid selection of above, or a fragment of any of
 CC these; preparation of a cell line that stably expresses an expression
 CC vector. The angiogenesis inhibitor has the following activities:
 CC antiangiogenic, angiogenesis stimulator, angiogenesis inhibitor,
 CC cytostatic, dermatological, antiarthritic, antirheumatic,
 CC antiinflammatory, vasotropic, hypotensive, ophthalmological,
 CC antiporiatic, and cardiant. The novel nucleic acid molecules of the
 CC invention may be used to treat disorders in gene therapy and antisense
 CC gene therapy. The nucleic acid sequences, proteins and antibodies as part
 CC of the therapeutic compositions are useful in treating a disorder of
 CC angiogenesis chosen from: tumour vascularisation, retinopathies,
 CC rheumatoid arthritis, Crohn's disease, atherosclerosis, ovarian
 CC hyperstimulation, psoriasis, endometriosis associated with
 CC neovascularisation, restenosis due to balloon angioplasty, tissue
 CC granulation due to scarification, peripheral vascular disorders,
 CC hypertension, vascular inflammation, Raynaud's disease, aneurism,
 CC arterial restenosis, thrombophlebitis, lymphadenopathy, lymphedema,
 CC ischaemia, angina, myocardial infarction, chronic heart disease,
 CC congestive heart disease, macular degeneration linked to age and
 CC osteoporosis. This polynucleotide represents a human DNA sequence for the
 CC creation of an angiogenesis inhibitor of the invention.

XX
 SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000118 Length: 7764
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conserv: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: 12 Gaps: 37

US-10-018-929C-3_COPY_650_1270 (1-621) x ADP73081 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
 DB 3342 GCGGACACAGCTGTCATCTTGTGACTCTGACTGGNACCCCGAATGACTTGCAGGCACAA 3401
 QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 DB 3402 GCCCGACGCATAGATTGTCAGAAAGACGAGGTAATATTACCGCTTAGTTACAAAG 3461
 QY 42 CysThrValGluGluLysAla-----Leu 49
 DB 3462 GGGACTGTGGAGGAGGAGATCATAGACGGCCCAAGAGAGATGGTATTAGATCATCTG 3521
 QY 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
 DB 3522 GTGATTCAGCGCATGGACACCACTGGCGGCAGCGATCTCTGGAAACAACACTCAGGAGGTCC 3581
 QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 DB 3582 AACTCAAATCCTTTAATAAAGAGAGAGCTACAGCTATTTTGAATTTGGAGCAGAGGAT 3641
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 3642 CTCTTCAAGAACTGGAA-----GGGAGGAATCA 3671
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120

DB 3672 GAACCTCAGGAATGAT-----ATAGATGAATTTTGGCGTGGCTGAA 3716
 QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 DB 3717 ACGAGAGAGAAATGAAGTGTCAACAAGTCAACAGAGATGAACCTTCTA---TCACAGTTTAAAG 3773
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 DB 3774 GTTGCCAACTTTCCAAATATGGAAAGATGAAGAGAGCTA-----GAAGAGCGT 3821
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 DB 3822 CCTCACAAGGACTGGGATGAGATCAATT-----3848
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 DB 3849 -----CCAGAGGAACAAGGAAAAAAGTAGAGGAGGAGCGGAGAGGAGCTAGAA 3902
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 DB 3903 GAAATTTATGCTGCTGCTGAATTCGAGTTCACCTAAAGAGCTCAGACAAATGACAGT 3962
 QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
 DB 3963 GACTCTGACACTGAGTCTAAGAGCGAGCCACAGATCCTCTGCTCTTGAGAGTGAAACG 4022
 QY 224 ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 DB 4023 GAAGAC-----TCTGATGATGACAAG--4046
 QY 244 LeuGluSerProLysValIleThrLeuGlnSerCysLysSerSerGlyThrAspGly 263
 DB 4047 -----CCAAAGCGCAGAGCGCTCCGAGAGTGTCGGAGAGACCTCTGTTGGAGGA 4097
 QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 DB 4098 TTTTACTGATGACAGATCCGAAGGTTTCATCAAGGCTTATAAGAAAGTTT-----4145
 QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 DB 4146 -----GGTCTCCCTCTTGAACCGCTGGAGTGTGTAGCAGCTGAT---GCTGAGCTG 4193
 QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 DB 4194 GTAGATAAGTCGGTGCAGATCTG-----RAGCGCTGGGT 4229
 QY 317 GlnValHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 DB 4230 GAACCTGATCCAC-----AACAGCTGTGTGTCTAGCAATGCGGAATATGAAGAGCAGCTG 4283
 QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
 DB 4284 AAAGAAAATGCCACGCGAGGAGGAAAGGACCCAGGAAAAGG---AGAGGTCCAACAATCAAG 4340
 QY 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 DB 4341 ATATCGGAGTT---CAGGTTAATGTGAATTCCTATTATCCACATGAAGAGGAGTTT---4394
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 DB 4395 -----CAGATGCTGCATAAATCTATCCCTGTGGACCTGAAGAAAAA 4442
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 DB 4443 TACTGCTTAACCTGTCGTGTCAAAAGCTGCACATTTTGTAGTAGAGTGGGGTGGAGAT 4502
 QY 415 AspCys-----PheGly-ThrAsnSer-----LY 422
 DB 4503 GATTCTCGCTGTGTGGGGATTTTGAACATGGCTATGGAACCTGGGAGTTAATATA 4562
 QY 422 sGlnSerValValSerThrLysLeuValAsn-----432

QY 203 yAspGlyGlyAsnAlaLysArgLysLysAlaSerAspAspValThrAspProArgVa 223
 Db 1478 CCATGGCTCGACTTCGGTTTTCGACCGGATCTCGGAGCACCTGCTGGATGTGATG 1537
 QY 223 lThrAspPro-----ProValAspAs 230
 Db 1538 GCITTTCCCAAGTCTCGGGTGGAGCTTCTCGCGAGCCCGCCAGCAACCACTGGGGAATAGAGA 1597
 QY 230 pAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer----- 246
 Db 1598 CAAGGATGACAGCGCAGTCCAGCAAGATGAGCTGCAGAGCAAGCAGTCCAAAGGCTTGA 1657
 QY 247 -----ProLysValIleThrLeuGlnSerSerCy 256
 Db 1658 GGAGAGGTACCATAGTTATCTCTCCACTTCCACACAGAGAGCGGGCCAGAGTCCCC 1717
 QY 256 sLysSer-----SerGlyThrAspGlyThrLeuAsp----- 266
 Db 1718 TCGCAGCTGGCCACTCAAGAGAGCCCGCCAGGGCCCGAGGGGCGCGAGTGGAA 1777
 QY 267 -----GlyAsnAspAlaPheGlyLeuTy-SerMetGlySerHisIleSe 281
 Db 1778 GGAGGACAGAGAGCTTGGGGAGGACTCTGCAGCCAGCCTCAGCCTGCAGCTGCTCCCTCCA 1837
 QY 281 r-GlyIleProGluAspMetLeuAlaSerGln----- 291
 Db 1838 GAGGCGATCCACAGAGCTGTGGCTCCCGCCAGAGCAGCTCTCAGAGGCTGCACATAAGGC 1897
 QY 292 -----AspTyrGlyLysIleProAspGluSerGlnArgArgLeuHisThrVall 308
 Db 1898 CATGGAAGAGCAGTGGCCCAAGTACTC-GAGCAAGACACAGAGCGCCTCTGCGAATCCA 1956
 QY 308 euLysProLysMet-----AlaLysLeuCySgln-----Vall 319
 Db 1957 AGCAAGAGAAGATGCAGCAACTGGCGGAGAGCTGTGCCAAGAGAGAGAGAGAGATCC 2016
 QY 319 euHisLeu-----SerAspAlaCyThrSerMetValGlyAsnPheLeuGluTyrv 336
 Db 2017 TCGGGCTTACCAGCAGCAAGAGAGCAATCTCTCAGTCTCTTGAGGAGCGGCTCGAAGAAG 2076
 QY 336 allLeuGluAsnHisArgIleTyrgluProAlaThrThrPheGlnAlaPheGlnIleA 356
 Db 2077 CCATTGAG-----GAGGAGGAGCGCCGATGAGAGAGAGAGAGAGAGAGAGAG 2121
 QY 356 laLeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuV 376
 Db 2122 GGTATCTCGCTCCGAGCT---CAGGTCCAGTCCAGCACACAGCAGATGAGGACCAAA 2178
 QY 376 alArgAlaAsnSerGluLeuAlaPheLysCySserArgValGluValAspTyriLeTyrs 396
 Db 2179 TCAGGGCTGAGCAAGAGGCTTCCCTGAGAAACTGAGAGAGAGAGTGGAG----- 2228
 QY 396 erIleLeuSerCyMetLys-----SerLeuPheLeuGluHisThrGlnGlyL 412
 Db 2229 -----TCTCACAGAGGCTGAGAGGCGCAGCTTGGACACAGAAAAATAGGCAATGC 2280
 QY 412 eu-----GlnPheAspCysPheGlyThrAsnSerLysGlnSerValVall 427
 Db 2281 TGGAGCAGCTCAAGGAAGAGATAGAGGCT---TCGGAGAAGAGCGAGCGCTGCCCTGA 2337
 QY 427 erThr-----LysLeuValAsnGluSerLeuSerGlyAlaThrValA 441
 Db 2338 ATGCTGCAAAAGGAGAGGCTCTGCAGCAGCTGGGGAGCAGCTGGAAGGGGAG-----A 2391
 QY 441 rAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMetT 461
 Db 2392 GGAAGAGAGCTGTGGCAACGCTGGAGAGGAGCAGCAGTGTGAG-----C 2436
 QY 461 hrGluLysArgCySerHisTy-SerThrAlaThrArgAspIle-----GluL 477
 Db 2437 TGGAGCGGCTGTGCTCTCTCATTTGGAGGCGCAAGCACCGGAGGTGCTTCCAGCCTCCAGA 2496

QY 477 ySThrIleSerGlyIleLysLysLysGlnValGlnLysLeuValGlnGluH 497
 Db 2497 AGAAGATACAGGAGCTCACAGAAAGAGGAGGCCCGCCAGCTGCAGAGTGCCTTGGGCAAG 2556
 QY 497 isGluGluLys-----LysMetGluLeuLeuA 506
 Db 2557 TGGAGCACAGAGTTCCACAGAGTCTTATCACGTGGTGGGTATGAGCAGCAGCTCAGCA 2616
 QY 506 snMetTyralaAspLysLysGlnLysLeuGlu----- 516
 Db 2617 GTCTCTCGAGAGAGAGCCAGGAGTGGAGGGGAGCAGATGAGAGGAGGTGGACAAGA 2676
 QY 517 -----ThrSerLysSerValGluAlaValI 526
 Db 2677 TGAAGAGAGGACACCAAGTATGCTAAGCCAGAGCAGTATGAAGCTGAGAGA 2736
 QY 526 leArgIleThrCySerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisA 546
 Db 2737 GGAAGCAGCGGGCTGAGCTTCTGGGCGCACCTGACCGAGAGCTGGAGCGCTCGAGAGGG 2796
 QY 546 snTyrgluArgLysPheAspGluIleLysSerGluLysAsnGluCySLeuLysSerLeuG 566
 Db 2797 CCATGAACGAGAACTGGAGACTGTGAGCAGAGCAACACAAAGCCTCTTGAGGACTGC 2856
 QY 566 luGlnMetHisGluValAlaLysLysLeuAlaGluAspGluAlaCySTrpIleAsnA 586
 Db 2857 GGCCTGGCCACAGGAGCAGGAGAAAGAGCTCCAGGATTTAGAGTTGGACCTTGAACCA 2916
 QY 586 rgIleLysSerTrpAlaAlaLysLeu 594
 Db 2917 GAGCTAAAGATGTCAAGGCCAGATTTG 2942
 RESULT 15
 ADS48467
 ID ADS48467 standard; cDNA; 3045 BP.
 XX
 AC ADS48467;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polynucleotide #3210.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX
 OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 DR
 XX New recombinant DNA construct comprising a promoter positioned to provide
 XX for expression of a polynucleotide encoding a polypeptide from a
 XX microbial source, useful for producing plants with improved properties.


```
Db 1998 GCAGTTGCAGAGCATCATTACAAATACGAAGCGAAAGGAAAGCCTTGAGAAGGAGGT 2057
QY 479 eSerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluG1 499
Db 2058 C-----AAGTTCAATAAGACGAGATCAAGAAGTTGAAGCAGGAGATCACTGA 2105
QY 499 uLysLysMetGluLeuLeuAsn-----MetTyrAl 509
Db 2106 GAAGAAACCTCACTGGCAACAACAGCAGCAGAGAGAGATATGCTCAAGGAAAGCTACGA 2165
QY 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaValIleArgIleTh 529
Db 2166 AGAGCAAAATCAAAAATCTCAATCGGACCACAAAAGGCTGCTCGGAGCTGAAAGTCAA 2225
QY 529 rCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAspHisAs 546
Db 2226 ACACCAAAACGAGCTGACGCGAGCTCAGAAAGGACGCGGACTTGAAA-----GAAACCAA 2279
QY 546 nTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuG1 566
Db 2280 CCTCCTGCAGAGTTAGACACACTCAGGCAGCAGAGATGAAGTCAGAGGAAACAGGCTTCA 2339
QY 566 uGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
Db 2340 AGCCGATTACGCGCTGAGAAAGCGAAGCTCAGCAAGGAT 2379
```

Search completed: July 11, 2005, 13:11:20
Job time : 940 secs

Db 2760 GCGGACACAGTCGTCATCTTTGACTCTGACTGGAACCCCGAAGTGCAGGACAA 2819
 QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLysSerVal 41
 Db 2820 GCCCGAGCGCATAGATTGGTCAGAAAGACGAGTAAATATTACCGCTTAGTTACAAAG 2879
 QY 42 CysThrValGluGluLysAla-----Leu 49
 Db 2880 GGGACTGTGGAGGAGGAGCATATAGAACGGGCCAAAAGAGATGGTATTAGATCATCTG 2939
 QY 50 IleLeuAlaArgGlnAnMetArgGlnAnLysAlaValGluAn-----64
 Db 2940 GTGATTGAGCGCATGGACACCACTGGCGGACGATCTCGAAACAACTCAGGAAGTCC 2999
 QY 65 -----LeuAnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 3000 AACTCAAAATCTTTTATAAAGAGAGCTGACACGCTATTTTGAAATTTGGAGCAGAGGAT 3059
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 3060 CTCTTCAAGAGATGGAA-----GGGAGGAATCA 3089
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
 Db 3090 GAACCTCAGGAATGGAT-----ATAGATGAAATTTTGGCGTTGGCTGAA 3134
 QY 121 GlyGluGlnAnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 3135 ACAGAGAGAGAAAGTGTCAACAAGTGCAACAGATGAATCTCTA---TCACAGTTTAAG 3191
 QY 141 TyrSerSerAspSerLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 Db 3192 GTTGCCCACTTTGCAACATGGAGATGAAGAGAGCTA-----GAAGACGT 3239
 QY 161 ProAnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 Db 3240 CCTCACAGGACTGGGATGAGATCATI-----3266
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 Db 3267 -----CCAGAGAACAAAGGAAAAAGTAGAGGAGGAAGCGGCAGAGGAGCTAGAA 3320
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3321 GAAATTTATATGTCCTCGAATTCGAGGTTCCTAAAGAGCTCAGACAAATGACAGT 3380
 QY 204 AspGlyGlyAsnAlaLysLysArgLysAlaSerAspValThrAspProArgVal 223
 Db 3381 GACTCTGACACTGAGTCTAAGAGCGCAGCCAGAGATCCTCTGCTCTGAGAGTGAAACG 3440
 QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 Db 3441 GAAGAC-----TCTGATGATGACAAGAG-----3464
 QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 Db 3465 -----CCAAAGCCAGAGCGCTCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 3515
 QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 Db 3516 TTCTACTGATCAGAGATCCGAAGTTTCATCAAGGCTTATAAGAGTTT-----3563
 QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 Db 3564 -----GGTCTCCCTCTTGAACCGCTGAGTGCATAGCAGCGTGAT---GCTGAGCTG 3611
 QY 297 ProAspGluSerGlnArgGluHisThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 3612 GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 3647
 QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 Db 3648 GAACTGATCCAC-----AACAGCTGTGTGTCAGCAATGCAGGAATACGAAGACGACGCTG 3701

QY 337 IleGluAnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
 Db 3702 AAGAATAATGCCAGCAGGGAAGAACGACGAGGAAAAGG---AGAGTCCAAACATCAAG 3758
 QY 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 3759 ATATCCGGAGTT---CAGGTTAATGTGAAATCCATTATCCAAATCAAGAGGAGTTT---3812
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 3813 -----CAGATGTCTGATAAATCTATCTCTGTGACCTCGAAGAAAAA 3860
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheGluHisThrGlnGlyLeuGlnPhe 414
 Db 3861 TACTGCTTAACCTGTCGTGTCAAGCTGCACATTTTGTATAGAGTGGGGGTGGAAGAT 3920
 QY 415 AspCys-----PheGly-ThrAsnSer-----Ly 422
 Db 3921 GATTCTCGCTGTGTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATTA 3980
 QY 422 sGlnSerValValSerThrLysLeuValAsn-----432
 Db 3981 ACAG-----ACCAGAGCTTAAATTAACACAAAAATCTCGCGGTGGAGACA 4028
 QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
 Db 4029 GATAAAAGCTCAGGGAGACGACGTACAGCCGAGCGGATTAATTGTTGAAGCTGCTC 4088
 QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
 Db 4089 AGAAGGGTCTGAGAGAAGAGGGGCTGTGACAGGTGGGAGACAGCCAAATTAAGAAG 4148
 QY 464 Arg-----464
 Db 4149 CGAAGCCTCGGTAAAGAGGAAAAACAAGTCCCGAGCTGAAAGAGGAGCATGGAATT 4208
 QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
 Db 4209 GAGCTTTTCATCTCTAGGCATTCAGATATCCATCAGAGGGGAGAGTGAAGATGAT 4268
 QY 481 GlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnHisGluGluLys 500
 Db 4269 GGCTTGGAAAAAGTCCAATGAAAAAACAACAAG---AAGAAAGAGAACAAAGGAGAAC 4325
 QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----511
 Db 4326 AAGGAGAAACAATGAGTTCTAGGAAGACAAAGAAGGGGACAAGGAAAGAAAGAGTCA 4385
 QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----523
 Db 4386 AAAGATAAGAAAGAGAGCCCTAAAGTGGTGTGTCGCCAATCTTCGAGTAATCAAGCGA 4445
 QY 524 -----AlaValIleArgIleThrCysSerArgThrThrGlnValGlyAsp- 539
 Db 4446 TCTCAGSGTCTCTCCATATTACAGCAGGAAGTGAACCTGTCTCCCATTTGGAGAGATGAG 4505
 QY 539 -----539
 Db 4506 GATGATGATCTGACCAGGAGACATTCAGCATATGTAAGGAGAGGATGAGCCCGTGAA 4565
 QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
 Db 4566 AAGGCACTGAACAGCTCGACAAACCTGCAAGGGGCTCAACGTGCAGAGACAGCTGGAA 4625
 QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer---564
 Db 4626 CACACCGGAACCTGCTGTCTGAAAAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCTAC 4685
 QY 564 -----564
 Db 4686 TCAGATCAGGAGCACATCAAACTCTGAGGAGGAACCTATGGATTTTGTTCGAAGTTT 4745

Qy 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
 Db 4746 ACAGAAATTGATCGAAACTGATTAAGTATATACCAAGATGGCTCATAGAAAGAGTCT 4805
 Qy 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
 Db 4806 CAAGAAGAGAG-----GAGCAAAAGAGAAAGACGAC 4838
 Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
 Db 4839 GTG-----ACTGGGGTAAAGAACCAATTTCGTCCAGAGGCTCAGGCTCCAGCGG 4889
 Qy 614 -----IleSerGlnAsn 617
 Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 2
 US-09-949-016-2989
 ; Sequence 2989, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2989
 ; LENGTH: 7699
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-2989

Alignment Scores:
 Pred. No.: 9,09e-08 Length: 7699
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conservative: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: 4 Gaps: 37

US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-949-016-2989 (1-7699)

Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
 Db 3343 GCGAGACAGTCGTCATCTTGACTCTGACTGGAAACCCCGAGATGACTTGCGAGCACAA 3402
 Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 Db 3403 GCCGCGCGCATAGAAATGGTCAGAGAGCAGCGTAAATATTTACCGCTTAGTTACAAAG 3462
 Qy 42 CysThrValGluGluLysAla-----Leu 49
 Db 3463 GGGACTGTGGAGGAGGAGATCATAGACGGGCCAAAAGAGATGGTATTAGATCATCTG 3522
 Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
 Db 3523 GTGATTGCGCGATGGAGACACCACTGGCGGAGCATCTCGGAAACAACACTCAGGAGGTCC 3582
 Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 3583 AACTCAATCTCTTTAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGAT 3642
 Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100

Db 3643 CTCTTCAAGNACTGGAA-----GGGAGGAATCA 3672
 Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
 Db 3673 GNACTCAGGAAATGGAT-----ATAGATGAAATTTTGGCGTGGCTGAA 3717
 Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 3718 ACGAGAGAGATGAAGTGTCAACAAGTGCACAGATGAATCTCTA---TCACAGTTTAA 3774
 Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 Db 3775 GTTCCCAACTTTCACAACATGGAGATGAAGAGACTA-----GAAGAGCGT 3822
 Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 Db 3823 CCTCAAGAGACTGGGATGAGATCATT-----3849
 Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 Db 3850 -----CCAGAGGAACAAGGNAAGAAAGTAGAGAGGAGGAGCGCAGAGAGGACTAGAA 3903
 Qy 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3904 GAATTTATATGCTGCTCGAATTCGGAGTTCCTAAAGGCTCAGACAAATGACAGT 3963
 Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
 Db 3964 GACTCTCAGACTGAGTCTAAGAGCGCAGGCTCTGCTCTCTGAGAGTGAACG 4023
 Qy 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 Db 4024 GAAGAC-----TCTGATGATGACAGAAG-----4047
 Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 Db 4048 -----CCAAAGCCGACAGGCGCTCCGAGGAGTGTGCGAAGGAGCTCTGTCGAGGGA 4098
 Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 Db 4099 TTTACTGATGACAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAGTTT-----4146
 Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyValIle 296
 Db 4147 -----GGTCTCTCTTGAACCGCTGGAGTGCATACACGCTGAT---GCTGAGCTG 4194
 Qy 297 ProAspGluSerGlnArgArgLysThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 4195 GTAGATAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 4230
 Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 Db 4231 GAACGTATCCAC-----AACAGCTGTGTGCAGCAATGCAGGAATACGAAGACAGCTG 4284
 Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
 Db 4285 AAAGAAAATCCACGCGAGGGAAGGACCGAGGAAAAGG---AGAGGTCCCAACATCAAG 4341
 Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 4342 ATATCCCGAGT---CAGGTATATGTGAATTCCTATCAACATGAAGAGGAGTTT---4395
 Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 4396 -----GAGATGCTGCATAAATCTATCCCTGTGGACCTGAAGAAAAA 4443
 Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 Db 4444 TACTGCTTAACCTGCTCGTCAAAAGCTGCACATTTTGTATGATGAGTGGGGGTGGAAGAT 4503
 Qy 415 AspCys-----PheGly-ThrAsnSer-----Ly 422
 Db 4504 GATTCTCGCTGTTGCTGGGGATTTTATGAACATGGCTATGGAACACTGGGAGTTAATTA 4563


```
Db 3248 -AGGAGCTCCAGGCGCCCTGGCCAGAGTGAAGAGGAGCTGCCAGAGAACATGGCC 3306
Qy 235 -----AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal-----11 250
Db 3307 CTCAAGAAGATCCGGGAGCTGGAATCTCAGATCTCTGAATCTCCAGGAAGACCTGGAGTCT 3366
Qy 250 eThrLeuGlnSerSerCysLysSerGlyThrAspGlyThrLeuAspGlyAenAspAl 270
Db 3367 GAGCGTGCTTCCAGGAATAAAGCTGGAAGCAGAAACGGGACCTTGGGGAA-GAG----- 3420
Qy 270 aPheGlyLeuTyrrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSe 290
Db 3421 -----CTAGAGCGGCTGAATAACAGAGTGGAGGACACGCTGGATTCACAGCTGCCCA 3473
Qy 290 rGlnAspTrpGlyLysIleProAspGluSerGlnArgLeuHisThrValLeuLysPr 310
Db 3474 GCAGGAGCTCAGGTCAAAACGTCAGCAGGAGGTGAACATCTTGAAGAAGACCTTGGAGGA 3533
Qy 310 oLysMetAlaLysLeu-----CysGlnValLeuHisLeuSerAspAlaCysThrSerMe 328
Db 3534 GGAG---GCCAAGACCCACGAGGCGCCAGATCCAGGAGATGAGGCAAGCACTCACAGGC 3590
Qy 328 tValGlyAsnPheLeuGluTyrrValIleGluAsnHisArgIle-----TyrGluG 345
Db 3591 CGTGGAGAGCTCGCGGAGCGAGCTGGAGCAGCAACGCGGTGAAGCAACCTCGAGAA 3650
Qy 345 uProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuVa 365
Db 3651 GGCAAGACAGACTCTGGAGAACGAGCGGGGGAGCTGGCCCAACGAGGTGAAGTGCTGCT 3710
Qy 365 1-----LysGlnIleLeuSerHisLysGluSerLeuVal----- 376
Db 3711 GCAGGGCAAGGGGAGCTCGGAGCAACAGCGCAAGAAAGTGGAGCGGAGCTGCAGGAGCT 3770
Qy 377 -----ArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
Db 3771 GCAGGTCAAGTTCAACAGGAGGAGCGCGTCCGACAGAGCTGGCCCAAGAGTGCACCAA 3830
Qy 388 g-----ValGluValAspTyrrIleTyrrSerIleLeuSerCysMetLysSerLeuPhe-- 405
Db 3831 GCTGAGGTGGAGCTGGACAACGTGACCGGCTCTCAGCCAGTCCGACAGCAAGTCCAG 3890
Qy 406 -----LeuGluHisThrGlnGlyLeuG 413
Db 3891 CAAGCTCACCAGGACTTCTCCGGCTGGAGTCCAGCTGCAGGACACTCAGGAGCTGCT 3950
Qy 413 nPheAspCysPheGlyThrAenSerLysGlnSerValValSerThrLysLeuValaenG 433
Db 3951 GCAGGAG-----GAGAACCGGCAGAGCTGAGCTGAGCACCAGCTC----- 3993
Qy 433 uSerLeuSerGlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAenSe 453
Db 3994 -----AAGCAGGTGGAGGACGAGAGAAATTCCTTCGGGAGCAGCTGGAGGA 4040
Qy 453 rSerGluAspGluCysMetThrGluLysArgCysSerHisTyrrSerThrAlaThrAr 473
Db 4041 GGAGGAGGAGGCCCAAGCAACCTGGAGAGACAGATCCGCCACCTCCATGCCAGGTGGC 4100
Qy 473 gAspIleGluLys-----ThrIleSerGlyIleL 483
Db 4101 CGACATGAATAAAGATGAGGAGCACTGTGGGGTGCTGTGAATACTGCTGAGGAGGTGA 4160
Qy 483 sLysLysTyrrLysLysGlnValGlnLysLeuValGlnGlnHisGluGluLys----- 500
Db 4161 GAGGAAGCTCCAGAGGACCTGGAGGCGCTGAGCCAGCGGCACGAGGAGGAGGTGGCGC 4220
Qy 501 -----LysMetGlu-----LeuLe 505
Db 4221 CTACGACAAGCTGGAGAGAACCAAGACGCGCTGCAGCAGGAGCTGACGACCTGCTGGT 4280
Qy 505 uAsnMet----- 507
Db 4281 GGACCTGGACCAACCGCCGAGAGCGGTGCACCTGGAGAGAGAGCAAGAGTTTGA 4340
```

```
Qy 508 -----TyrAlaAspLysLysGlnLysLe 515
Db 4341 CCAGCTCTGGCGAGGAGAGACCATCTCTGCAAGTATGCAGAGGCGCGACCGGC 4400
Qy 515 uGluThr-----SerLysSerValGluAlaAlaValIleArgIleThrCysSerAr 532
Db 4401 TGAGCGGAGCGCGAGAGAGAGACCAAGGCTCTCTGCTGGCCCGGCGCTGGAGGA 4460
Qy 532 gThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAenTyrrGluArgLysPheA 552
Db 4461 AGCATGGAGCAGAGCGGAGCTGGAGCGCTCAACAGCAGTTCGCACGAGAGATGA 4520
Qy 552 pGluIleLysSerGlnLysAenGluCysLeuLysSerLeuGluGlnMetHisGluValAl 572
Db 4521 GGACCTTATGAGCTCCAAAGATGATGTGGCAAGAGT-----GTCCACGAGCTGA 4571
Qy 572 aLysLysLeuLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAl 592
Db 4572 GAAGTCCAAAGCGGCGCTAGAGCAGCAGG--TGGAGGAGATGAAGACGCACTGGAAGAG 4629
Qy 592 aLysLeuLysValCysValPro 599
Db 4630 CTGAGGACGAGCTGCAGGCCA 4651
```

RESULT 4

US-09-220-132-79

; Sequence 79, Application US/09220132

; Patent No. 6506607

; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 07334-074001

; CURRENT APPLICATION NUMBER: US/09/220,132

; CURRENT FILING DATE: 1998-12-23

; PRIOR FILING DATE: 1998-03-25

; PRIOR APPLICATION NUMBER: US 60/079,303

; PRIOR FILING DATE: 1997-12-24

; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 79

; LENGTH: 5857

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-220-132-79

Alignment Scores:

Pred. No.: 8.59e-07 Length: 5857

Score: 156.00 Matches: 136

Percent Similarity: 39.80% Conservative: 106

Best Local Similarity: 22.37% Mismatches: 212

Query Match: 4.85% Indels: 155

DB: 4 Gaps: 28

US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-220-132-79 (1-5857)

Qy 31 GluArgThrLysIlePheArgLeuTyrrSerValCysThrValGluGluLysAlaLeuIle 50

Db 1840 GAAGAAACTCATCAGAAAGGAGATAAAGGCTCTGTATACCGCCCGCAAGAAAG----- 1890

Qy 51 LeuAlaArgGlnAenMetArgGlnAsnLysAlaValGluAenLeuAenArgSerLeuThr 70

Db 1891 CTTTCCAAAGAGAACGAGTCATTGAAAAGCAAGCTGGAGCATGCCAACAAAGAGAACTCA 1950

Qy 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrrLeuPheAspLysLeuAspHisPheHiser 90

Db 1951 GATGTGATAGCTCTATGGAAGTCC-----AAACTG----- 1980

Qy 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110

Db 1981 -----GAGACTGCCATCCACAGCAGCGGATGGAAGAACTGAAGGTA 2028

```
QY 111 GluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeu----- 128
Db 2029 TCTTTGAGCAAGGGCTTGGAAAGCAGAGACGGCAGAAATTTCTGCACTTAAACACACAAATA 2088
QY 129 ---CysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSer-AspSerThrLe 147
Db 2089 GAGAAATGAGACTAGATTACCAACACGAATAGAAAATTTGCAGAACTCAACAAGACTCT 2148
QY 147 uPheGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSer-- 166
Db 2149 GAACGGGCTCCCATGCTAAAGAGA-----TGAAGCC 2181
QY 167 -----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrProS 180
Db 2182 TTGAGGGCTAAACTGATGAAGTTATTAAAGAAAGGAAACACAGCTCGGAAGCCATCAGG 2241
QY 180 exAspThrProGln-----ArgAsnAArgLysArgValGlnTyrP 193
Db 2242 TCGAACTGCACAAAGCAGAGACCAGCATCTCGTAGAATGGAGACAGCTTAAACAAA 2301
QY 193 heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys 210
Db 2302 TTACAGGAAGCTGAATAAAGGTAAGGAGCTAGAGTACTGCAAGGCCAAATGCAATGAA 2361
QY 211 ArgLysLysAlaSerAspAspValThrAspPro---ArgValThrAspProProValAsp 229
Db 2362 CAAACCAAGGTTATTGATAATTTTACATCAGCTCAAGGCTACTGAAAGAAAGCTCTTG 2421
QY 230 AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu 245
Db 2422 GATCTTGATGCATTCGGAAGCCAGTTCGAGAGTTAAATCGGAATGAGAACTTAGA 2481
QY 246 SerPro-----LysValIleThrLeuGln-----Ser 254
Db 2482 CAGCAGCTTGAGGCAGCTGAGAAACAGATTAAACATTTAGAGATTGAAAGAAATGCTGAA 2541
QY 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
Db 2542 AGTAGCAAGCTAGTACATTTACCAGAGAGCTCCAGGGGAGAGAG-----CTAAAGCTTACT 2598
QY 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly 294
Db 2599 AACCTCAGGAAATTTGAGT-----GAGCT-CAGTCAAGTGAAGAGACTTTGGA 2648
QY 295 LysIleProAspGluSerGlnArgArgLeu-----HisThrValLeu 308
Db 2649 AAAAGAACTTCAGATTTTGAAGAAAGATTTGCTGAAAGCTTCAGAGGAGGAGCTCTCTGT 2708
QY 309 LysProLysMetAlaLysLeuCysGln-ValLeuHisLysSerAspAlaCysThrSerMe 328
Db 2709 TCAGAGAGAGTATGCAGAAACTGTAAATAAGTTACACCAAAAGGAGGAGCAAGTTTAACAT 2768
QY 328 tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaTh 348
Db 2769 GCTGCTCTTGACTTGAGAGCTGAGAGAAAATCTTAGCAGATATGAGGGCAAAA----- 2823
QY 348 rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuValLysGlnI 368
Db 2824 -----TTTAGAGAGAA 2834
QY 368 eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
Db 2835 AGATGAGAGAGAACAGCAGCTGATAAAGCGCAAGGAAAGAACTG----- 2877
QY 388 gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi 408
Db 2878 -----ATTGCAGAAATATGAG----- 2904
QY 408 sThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerTh 428
Db 2905 -----ATGTCAGGAGATAACTCTTCTCAG-----CTGAC 2933
```

```
QY 428 rLysLeuValAsnGlu---SerLeuSerGlyAlaThrValArgAspGluLysIleAsn-- 446
Db 2934 AAAATGACAGCATGATTAATGCTGAAACAAAGAGATGTAGAAGATTACAGCTAAAACT 2993
QY 447 -ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSe 466
Db 2994 TACAAAGGCTAATGAAAATGCAAGT----- 3018
QY 466 rHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTy 486
Db 3019 -----TTTCTGCAAAAAGTATTGAGGACATGACTGTCAAAGC 3056
QY 486 rLysLysGlnValGlnLysLeuValGlnHisGluGluLysLysMetGluLeuLeuAs 506
Db 3057 TGAACAGAGCCAGCAAGACAGCTAAAAAGCATGAGGAAGAAAGAAAGAAATTTGGAGAG 3116
QY 506 nMetTyrAlaAspLysGlnLysLeuGluThrSer-----LysSerValG 522
Db 3117 GAAATTTGTCGACCTGGAAGAAAATGGAACCAACAGCCACACAGCTGTGACGAGCTGAA 3176
QY 522 uAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu---- 540
Db 3177 AGCCAGGTATGAGAGAGCCACT---TCTGAGACAAAACCAAGCATGAAGAATCTCTACA 3233
QY 541 -----LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSe 556
Db 3234 GAACCTCCAGNAGACGCTGCTGGAC-----ACAGAGGACAAAGCTGAAGGGCCACGGGA 3287
QY 556 rGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLe 576
Db 3288 GGAGAACAGTGGCTTCTGTCAGGAGCTGGAGGAGCTGAGAAAGCAAGCCGACAAAGCCAA 3347
QY 576 uAlaGluAspGluAla 581
Db 3348 AGCTGCTCAACAGCG 3363
RESULT 5
US-09-248-796A-6172
; Sequence 6172, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6172
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-6172
Alignment Scores:
Pred. No.: 9,14e-07 Length: 1893
Score: 148.50 Matches: 133
Percent Similarity: 37.46% Conservative: 97
Best Local Similarity: 21.66% Mismatches: 222
Query Match: 4.62% Indels: 162
DB: 4 Gaps: 27
US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-248-796A-6172 (1-1893)
QY 95 AspSerGly-ValSerPheGluGlnSerIleMet-----AspGlyValIleHi 110
Db 66 GATTTGGAAGTTGAAAAGAAAGAGAGCTGCTAGTGAATAATCTGAAACTGTCTATAA 125
QY 110 sGluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeuCysLe 130
```



```

126 GGAATACAGTGACAAAATAAGAGTTTGAAGTAAGATCAACTCAATAAAG----- 177
130 uLeuLeuGluAlaLysHisAlaGln-----GlyThrTySerSerAspSerThrLeuPh 148
178 -----GAGAAATCATTTCAAAAGAGATAAACAACATCAATGAACAAAAGACTTCTTT 227
148 eGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLe 168
228 GAAACAAGATATTGCAAAACTATCTCAAGATCAT-----GAACTGGCGCAAACTCAATT 281
168 uLeuGlyGlyLysAsnProMetTrpLysTyProSerAspThrProGlnArgAsnArgLys 188
282 AGAAGATAAGAGAGATCAATTTGAAGGAATGAAAGCTAGTTTAGAGAAACAACAAT----- 336
188 sArgValGlnTyPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAl 208
337 -----ACTGAGTCGCAACCAACGATTCGAGAAAGAACAAAT-- 372
208 aLysLysArgLysLysAlaSerAspValThrAspProArgValThrAspProProVa 228
373 ----CAGATAAGGAAGACTCTCGAAACTATTAAAGTCAATTAATAA--ACTGAATTAAGAC 425
228 lAspAspGluArgLysAlaSerGlyLysAspHis----- 240
426 AAGTGGTGATGCATTAAACAATCTCAAAAGGAATACAAAACCTTTAAAACTAAAAATTC 485
241 -----MetGlyAlaLeuGluSerProLysVal----- 249
486 CGACACTGAACTGAATTTGGAAAACAATTTGGAAGAACTAGAAAAGTGAAGCTGGATT 545
250 -----lLeuThrLeuGlnSer-- 254
546 ACAAACAGCAGCATGAAAATTTGAAAGGAATTTACAGAAAGGGAATTTGCATTTGAAATCTGA 605
255 -----SerCysLysSerSerGlyThrAspGlyThr----- 264
606 ACTTGAACAGTGAATAAATAGTGGATTGTTTACAACTCTCGAGCTTGTGCAATTCGACAAA 665
265 -----LeuAspGlyAsnAspAl 270
666 AACAGTGAATCATTTGAAAAGAAAGAAAGGAAGATTTGCAATTTTATCGGCAACAATC 725
270 aPheGlyLeuTySerMetGlySerHisIleSerGlyLeProGluAspMetLeuAlaSe 290
726 AAAGCAATTGCAAGATTACATTTCAAAAGCAGATGATATTTTCGGAAGAACTTAAAGGCATT 785
290 rGlnAspTrpGlyLysIlePro-----AspGluSerGlnArgArgLeuHisTh 306
786 AACTGATGAGCTTAAAGAGAGAGACCAACACAGCTTTGATGATTTCAAAAGAAATTA--AC 842
306 rValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAlaCysTh 326
843 TGAGTTGAAATGATTTGAGCTTCCACTAAGAGGAATTTGAAACTGAAAGACACAGAC 902
326 rSerMetValGlyAsnPhe-----LeuGluTyrValIleGluAsnHisAr 341
903 GAGTAAATTTCAAAACCTCGAGGAAGAGAAAGAAAGAAATTTGAAACTCAATAAGA 962
341 gIleTyGluGluProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAl 361
963 ATTAGAGTTATTGAAAATGCAATAGTGGTCTTAAAGAAAGGAAGAACTTCTGGAAGATATC 1022
361 aAlaLeu-----LeuValLysGlnIleLeuSerHisLysGluSerLeuVa 376
1023 CAAGCTTGAGTCAGAAATTTGAGATTTTATCCAAAGAGTTAGAACACAAAATTCGGTCAT 1082
376 lArgAlaAsnSerGluLeuAla-----PheLysCysSerArgVa 389
1083 GAAACAACATGATGAATTTGAAAGAAACAACAAGGAAAAAATCAAGAGCTACAAAAGT 1142
389 lGluValAspTyTrpIleTySerIleLeuSerCysMetLysSerLeuPheLeuGluHisTh 409

```

```

1143 TACAAGAGGATTTCTACTACCAAACTGAAA---TTAGATGAATTCGAGAAAGAAATTAGA 1199
409 rGlnGlyLeuGlnPhe---AspCysPheGlyThrAsnSer-----LysGlnSerVa 425
1200 TGCTGCATTTGTCATTTCAAGATAGATTTGAAACACAGCTTCAGCCAAAGTTAACACAATCTAC 1259
425 lValSerThrLysLeuValAsnGluSerLysLeuSerGlyAlaThrValArgAspGluLysIle 445
1260 TTCTGATTGGAGCAGCAAAATAAGAAATTTGAATTTTGTATTAGTGAGAAAGAAAGAAAC 1319
445 eAsn-----ThrLysSerMetArgAsnSerSerGluAspGluCysMetThrG1 462
1320 TGAACAAGAACTTCAAAAACCTTACAAAACAACGCTGAACTTGAAGAACTGATTGATGA 1379
462 uLysArgCysSerHisTySerThrAlaThrArgAspIleGluLysThrIleSerGlyI1 482
1380 TAAAATGCT-----GATTCAGTAAGACAAAATTTCTGATCT 1415
482 eLysLysLysTyLysLys-----GlnValGlnLysLeuValGlnG1 496
1416 TGAAGAAAGCCCAACAGAAAGATTTGGACCTTGAACCGCAACTTGAAGAACTTACCAAGA 1475
496 uHisGluGluLysLysMetGluLeuLeuAsnMetTyAlaAspLysLysGlnLysLeuG1 516
1476 ACTTGATCTAATAAACTGGAATTA-----CAAAAACATGAAAATTTG-- 1518
516 uThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerG1 536
1519 -----AACAAATGA 1526
536 nValGlyAspLeuLysLeuAspHisAsnTyGluArgLysPheAspGluIleLysSe 556
1527 ATTCACCAATTTGATGTAATTCACATGAAGAAATTTAAAACTTAAATTAGACAAAATTTGAAA 1586
556 rGluLysAsnGluCysLysSerLeuGluGlnMet-----HisGluValAlaLys 573
1587 TGAAGAGATGATCTTGCAATCATCATGAAGAACTCAATGTTGCAACATAAGAAACGTC 1646
573 sLysLysLeu-----AlaGluAspGlu-----AlaCysTrpIleAsnAr 586
1647 AACAAATTTCAAAATGTTGAAGAAAGAAATAAGAAAATTCACAGCAAAATATGAATCTGA 1706
586 gIleLysSerTrpAlaLysLysLeuValCysValPro 599
1707 ACTCAAGAAATTTACAAGATAAACTTAAAGAAATTTGTGTTCT 1746

RESULT 6
US-09-248-796A-183
; Sequence 183, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 183
; LENGTH: 4947
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-183

Alignment Scores:
Pred. No.: 6.48e-06 Length: 4947
Score: 147.50 Matches: 133
Percent Similarity: 36.72% Conservatve: 102
Best Local Similarity: 20.78% Mismatches: 243

```

Query Match: 4.59% Indels: 163
DB: 4 Gaps: 28

US-10-018-929c-3_COPY_650_1270 (1-621) x US-09-248-796A-183 (1-4947)

60 LysAlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSer 79
DB 2698 AAAAATTTGGACGATTTGAATAAATCATTCGAAGAGAAATTTGAATTTGGCAACACTC 2757
QY 80 TyrLeuPhe-----AspLysLeuAspHisPheHisSerSer 91
DB 2758 TTTTATTTTATTCAAACCAATCCCATTTATTTGACAAAATTTG-----TTTAATTCAT 2811
QY 92 GluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHisGlu 111
DB 2812 CCATATACATAAAATCAGACAAATCTGGTCAGATTTTGGTCAGTCTGTGATCAATTA 2871
QY 112 Phe-----SerSerIleLeuSerSerLysGlyGlyGluGluAenGluValLys 127
DB 2872 TTCCAGTAGAGATTTCATCAATTTACTTATCATCTCTCGTGAAGAGTACTTTTGGTCAA 2931
QY 128 LeuCysLeu---LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThr 146
DB 2932 CTTATGATCCAGTTGATGCAAAATGATACAGCTAACTCAAAACAAATTTGGGTGATCACC 2991
QY 146 ----- 146

2992 AGTTGCAATTTGACAAACTCGATAGATTCTTTACTAAATTTCAATATCATATTCAG 3051
QY 147 -----LeuPheGlyGluAspHisIleLysLeuSerAsp----- 157
DB 3052 AGACAACTTTGAAGACACTTTTGGCAAGTTTGTATATAGTATTTGTGATAGAACAA 3111
QY 158 -----GluGluSerProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
DB 3112 GTGGATTTTCAATCAGATCCAAATCAGAATCTACAATCAATCAAT-----GATCATGAG 3165
QY 175 MetTrpLysTrpProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
DB 3166 ATGAAGTTTATGGTCTAGTGAAGCTCTCGTGATATTTCCACACAGGACGCTATTTCAG 3225
QY 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysAla 214
DB 3226 TTCCAGAACTTAGTAATAAGTTT-----GTGGCAATTTGATGCTATTAAAGGAGACT 3279
QY 215 SerAspAspValThrAspProArgValThrAspProProValAspAspGluArgLys 234
DB 3280 TGCAGTGATTTGCTTTCATATGTTA-----AspHisMet----- 241
QY 235 AlaSerGlyLys-----AspHisMet----- 241
DB 3313 GCTTCTGGAATAAAGCTTTGTGCAATCCCTGACCATGTGAAATGATTGTGCTGCA 3372
QY 242 GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThr 261
DB 3373 GGTATTATTGCTCTCAAGAAAGTTCCCGACAAATCTGATCAACAACTTTAGCGGTT 3432
QY 262 AspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSer 281
DB 3433 GCTGGGGTTATA-----TTTGTGAACACTATCTTGGATCT-...ATTG 3474

282 GlyIleProGluAsp-----MetLeuAlaSerGlnAspTrpGlyLysIleProAsp 298
DB 3475 CAAGTACCAAGACATAGGAATTTTGCAGTGAACAAATGATATCCAGAAAGCTAAAGC 3534
QY 299 ---GluSerGlnArgArgLeuHisThrVal----- 307
DB 3535 AAAGACAAATTTGGATACCTATCGTGTATGTTGCAATTTGTTTCAATGAACCACTTT 3594
QY 308 -----LeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAsp 323
DB 3595 AATGACAAATTTTGTGAACCACTTGAATGAG-----TACATTATGGCAAGTACCGAC 3645

QY 324 AlaCysThrSerMetValGlyAenPheLeuGluTyrValIleGluAenHisArgIleTyr 343
DB 3646 ACCGTTAAGTCAATCATTTAGTCAGGCAATTTAATGTTGGTGAATTTGAAACCGTGTAT 3705
QY 344 Glu-----GluProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrp 359
DB 3706 GAATTACATGATTTATGATCATCTGTGACTCATCAGAGACCTAAGTTGACTATCAGTGTA 3765
QY 360 IleAlaAlaLeu---LeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAla 378
DB 3766 AACTCATTTGATTCATTAAGAGAAATCCATCTTGCATAATGTCATATATTATCAACAGGC 3825
QY 379 AsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398
DB 3826 AATGATGATCAATTTGTACAAA---ACATCGTCGAGGTGGAAAAATG-----TTGATT 3876
QY 399 SerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGly 418
DB 3877 TCCCAACAAGACATGTTGACATTTGACCGATTTGTCATCGGTTACTTTGAACCTTAAATCCT 3936
QY 419 ThrAsnSerLysGlnSerValSerValSerThrLysLeuValAsnGluSerLeuSerGlyAla 438
DB 3937 ACCACAAGAGAGGTCTATAGTTGATTCAAAA----- 3969
QY 439 ThrValArgAspGluLysIleAenThrLysSerMetArgAsnSerSerGluAspGluGlu 458
DB 3970 -----ACTAAAACATTTGTTCAACCAAGCC----- 3993
QY 459 CysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp----- 474
DB 3994 -----AAACGCTGCTTGTGTATATCATTTAGATGTGCAGGAAGAAGACGACGT 4041
QY 475 -----IleGluLysThrIleSerGlyIleLysLysLysTyrLysLysGlnValGln 491
DB 4042 GATGATTTATTTGAATTTAATATATCGGCATCAAAACCAAGCCACCAAGATTTCAAG 4101
QY 492 LysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLys 511
DB 4102 GAGATTTTCAATATGAAAA-----GCGGAGCAA 4131
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIle-----ArgIle 528
DB 4132 GATATTTCCCTCAACCAACGACGAGGTGTTACTGCTTAATCTGCTCAATAAAGAAATCT 4191
QY 529 ThrCysSerArgThrSerGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGlu 548
DB 4192 ACCAGACCATATTCGGGTACATCATTTAGGGGACTTGTGCAATTTAACCTTATCAT- 4245
QY 549 ArgLysPheAspGluLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMet 568
DB 4246 -----GAATTTGAAAAAATGTTGTTGGAGATTATTCTTAAGCTTGAATCAATG 4293
QY 569 HisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLys 588
DB 4294 GGGAGTGTGACTCGTAAAAAATCATTTCAA-----ACATGCTCATCAATAGCT 4344
QY 589 SerTrpAlaAlaLysLysValCysValProIleGlnSerGlyAenLysHisPhe 608
DB 4345 A--TGGACATTAACAAAGGATTTCAAAGACAAATGCCGCTTGGCAACATTCGAAAGTTT 4402

RESULT 7
US-09-949-016-1054
; Sequence 1054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1054
 ; LENGTH: 6016
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-1054

Alignment Scores:

Pred. No.: 1,05e-05 Length: 6016
 Score: 147.00 Matches: 137
 Percent Similarity: 35.65% Conservative: 104
 Best Local Similarity: 20.27% Mismatches: 257
 Query Match: 4.57% Indels: 178
 DB: 4 Gaps: 27

US-10-018-929c-3_COPY_650_1270 (1-621) x US-09-949-016-1054 (1-6016)

Qy 19 LysHisValGluLysIleLeuSerCysSerGluArg---ThrLysIlePheArg 37
 Db 3025 AAACATGCCACAGACAGGTGAACCTCAGAGAGAGATGGCAGGTCTCGATGAA 3084
 Qy 38 LeuTyrSerValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArg 57
 Db 3085 ACCATTGCTAGCTGACCAAGGAGAGAGAGCTCTCCAGGAGGCCACCACGACGACCTG 3144
 Qy 58 GlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuMetTrpGly 77
 Db 3145 GATGACCTGCAGATGGAGGAG----- 3165
 Qy 78 AlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGly 97
 Db 3166 -----GACAAAGTCAACACCTGACCAAGCTAAAGCTAAAC----- 3198
 Qy 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
 Db 3199 ---AAGCTAGAACACAGAGTGGAGAT----- 3222
 Qy 118 SerLysGlyGluGluAsnGluValLysLeuCysLeuLeuLeuGlu-----AlaLys 135
 Db 3223 CTTGAGGATCTCTGGAACAGAAAGAACTTTGTCATGAGCTTAGAAGAGCCAAAGAGA 3282
 Qy 136 HisAlaGlnGly-----ThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHis 152
 Db 3283 AAATGGAGGTGACCTTAAATTTGGCCCAAGAAATCCACATGGATACAGAAATGACAAA 3342
 Qy 153 IleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuLysGlyLys 172
 Db 3343 CAGCACTTAATGAGAACTCAAAAGAAAGAGTTTGAATGAGCAATCTGCAAGGCAAG 3402
 Qy 173 AsnProMetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGln--- 191
 Db 3403 ATTGAAGATGAACAGCCCTTGCAATGCAGCTACAAAGAAAGATCAAGAAATACAGGCC 3462
 Qy 192 TyrPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArg 211
 Db 3463 CGCAATTGAGGAGCTGGAGAGGAGAAATCGAGCAGAGCGCGGCTCCCGGGCCAAAGCAGAA 3522
 Qy 212 LysLysAlaSerAsp-----AspValThrAsp----- 220
 Db 3523 AAGCAGCGCTCTGACCTCTCCCGGAGCTGGAGGAGATCAGTGAGGCTGGAAGAGGCC 3582
 Qy 221 ProArgValThrAspProProValAsp---AspAspGluArgLysAlaSerGlyLysAsp 239
 Db 3583 GGTGGGGCACTTCAGCCCAAGATTGAGTTGAACAGAAAGCGGAGGCTGAGTTCAGAAA 3642
 Qy 240 HisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSer 259
 Db 3643 ATGCGCAGGGACCTGGAA-----GAGTCCACCTCGACGACGAA 3681

Qy 260 GlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHis 279
 Db 3682 GCCACGCGCAGCTGCTCTTCGGAAGACGACGACATAGTGTGGCTGAGCTTGGGAAGCAG 3741
 Qy 280 IleSerGlyIlePro----- 284
 Db 3742 ATCGACAGCCTTCAGCGGGTCAAGCAGAGCTGGAGGAAGGAAAGAGTGAGCTGAAGATG 3801
 Qy 284 ----- 284
 Db 3802 GAGATCAATGACCTTGCTAGTAACATGAGACTGCTCCAAAGCCAAAGCAAACTTTCAG 3861
 Qy 285 -----GluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAsp 298
 Db 3862 AAAATGTGCGCAGCCCTAGAGGACGACCTTAGTGAATA-----AAAACAAGGAA 3912
 Qy 299 GluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu----- 315
 Db 3913 GAAGCAGCAACGCTTAATAATGAGTTGTGCCCAAGGAGGACGCTTACACACAGAA 3972
 Qy 316 -----CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsn 331
 Db 3973 TCAGGTGAGTTTCCACGACGCTAGATGAAAGATGCTATGTTTCTCAGCTATCCGA 4032
 Qy 332 -----PheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu----- 345
 Db 4033 GGCACAAACAGCATTTACACACAGATTGAAGATTAAGAGGCGACGTAGAGAGGAGACT 4092
 Qy 346 -----ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeu 363
 Db 4093 AAGCCCAAGGACCTGCGCCCATGCTGCGTGCAGTCAGCC---CGCCATGAGCTGTGACCTG 4149
 Qy 364 LeuValLysGlnIleLeuSerHisLysGluSer-----LeuVal 376
 Db 4150 CTGCGGGAACAGTATGAGGAGGAGCAGCAAGCCAGGCTGAGCTGCAGAGGGGAATGTC 4209
 Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSer 396
 Db 4210 AAGCCCAACAGTGAAGTTGCCAGTGAGGAGCAACGAATGACGAGCGGACGCCATC----- 4263
 Qy 397 IleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCys 416
 Db 4264 -----CAGCCACAGAGGAGCTGGAG----- 4284
 Qy 417 PheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSer 436
 Db 4285 -----GAGCCCAAGAGAGAGCTAGCCAGCGCTCTGCAG 4317
 Qy 437 GlyAlaThrValArgAspGluLysIleAsnThrLys-----Ser 449
 Db 4318 GATCGAAGAACATGTAGAAGCTGTGAATTCCAATGTGCTCTCTTGAAGACAAAG 4377
 Qy 450 MetArgAsnSerSerGluAspGluGluCysMet-----ThrGluLysArg 464
 Db 4378 CAGAGGTACAGATGAAGTAGAGGACCTCATGATGATGTGGAACAGCTTAATGCTGCC 4437
 Qy 465 CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLys 484
 Db 4438 TGCATAGCTCTCGATAGAGCAAGAAACTTTGACAAGGTTCTGGCAGATGGAAACAG 4497
 Qy 485 LysTyrLysLysGlnValGlnLeuValGlnGluHisGluGluLysLys----- 501
 Db 4498 AAGTATCAGGAAACTCAGGCTGAACTTGGAGCCCTCCAGAGGAGTGGCGTCTCTCAGC 4557
 Qy 502 -----MetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSer 518
 Db 4558 ACTGAGCTGTTCAGAGTGAAGATGCGCTACGAGGAATCCCTGGATCATCTTGAACCTTA 4617
 Qy 519 LysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
 Db 4618 AAGCGA---GAGATAAGAACTTACACAGGAGATTCTTGACCTGACAGCAAAATGCA 4674

```

Qy 539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluLeuLysSerGluLys 558
   ::      |||      |||      |||      |||      |||      |||      |||
Db 4675 GAGGGTGGAAAGCATATCCATGAACCTGGAGAAAGTAAAGAAACAACCTTCATCATGAGAAG 4734
   ::      |||      |||      |||      |||      |||      |||      |||
Qy 559 AsnGluCysLeuLysSerLeuGluClnMet-----HisGluValAlaLys 573
   ::      |||      |||      |||      |||      |||      |||      |||
Db 4735 AGTGAACCTACAGACTTCCTTAGAGGAAGCAGAGGCATCTCTTTGAGCATGAAGAGCCAAA 4794
   ::      |||      |||      |||      |||      |||      |||      |||
Qy 574 LysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSer 589
   ::      |||      |||      |||      |||      |||      |||      |||
Db 4795 ATTCTTCGCATTCACCTTGAG-----CTAAATCAGGTGAATCT 4833
   ::      |||      |||      |||      |||      |||      |||      |||

RESULT 8
US-09-949-016-2264
; Sequence 2264, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2264
; LENGTH: 6016
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2264

Alignment Scores:
Pred. No.:      1.05e-05      Length:      6016
Score:          147.00      Matches:      137
Percent Similarity: 35.65%      Conservative: 104
Best Local Similarity: 20.27%      Mismatches: 257
Query Match:      4.57%      Indels:      178
DB:              4          Gaps:      27

US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-949-016-2264 (1-6016)
Qy 19 LysHisValGluLysIleLysIleGluSerCysSerGluArg---ThrLysIlePheArg 37
   ::      |||      |||      |||      |||      |||      |||      |||
Db 3025 AAACATGCCACAGAGAACAGGTGAAAAACCTCAGAGAGAGATGCGAGGTCTGGATGAA 3084
   ::      |||      |||      |||      |||      |||      |||      |||
Qy 38 LeuTyrSerValCysThrValGluCylLysAlaLeuIleLeuAlaArgGlnAsnMetArg 57
   ::      |||      |||      |||      |||      |||      |||      |||
Db 3085 ACCATTGTAAGCTGACCAGGAGAGAAAGGCTCTCCAGGAGGCCCCACGAGACCCCTG 3144
   ::      |||      |||      |||      |||      |||      |||      |||
Qy 58 GlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGly 77
   ::      |||      |||      |||      |||      |||      |||      |||
Db 3145 GATGACCTGCAGATGGAGAG----- 3165
   ::      |||      |||      |||      |||      |||      |||      |||
Qy 78 AlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGly 97
   |||      |||      |||      |||      |||      |||      |||      |||
Db 3166 -----GACAAAGTCAACCCCTGACCAAAGCTAAAACC----- 3198
   |||      |||      |||      |||      |||      |||      |||      |||
Qy 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
   |||      |||      |||      |||      |||      |||      |||      |||
Db 3199 ---AAGCTTAGAACCAAGATGGACGAT----- 3222
   |||      |||      |||      |||      |||      |||      |||      |||
Qy 118 SerLysGlyGlyGluGluAsnGluValLysLeuCysLeuLeuGlu-----AlaLys 135
   |||      |||      |||      |||      |||      |||      |||      |||
Db 3223 CTTGAAGATCTCTGGAAACAGAAAGAAACCTTTGCATGCATCTTAGAAGACCAAGAGA 3282
   |||      |||      |||      |||      |||      |||      |||      |||
Qy 136 HisAlaGlnGly-----ThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHis 152
   |||      |||      |||      |||      |||      |||      |||      |||

```

```

Qy 437 GlyAlaThrValArgAspGluLysIleAsnThrLys-----Ser 449
Db 4318 GATCGAAGAAACATGTAGAGCTGTGAATTCCTTCAAAATGTCTCTTTGAAAGACAAAG 4377
Qy 450 MetArgAsnSerSerGluAspGluGluCysMet-----ThrGluLysArg 464
Db 4378 CAGAGCTACAGATGAAGTAGAGAGCTCATGATGTGTGGAAAGCATTAATCTGCC 4437
Qy 465 CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLys 484
Db 4438 TGCTAGCTCTCGATAAGACCAAGAACTTTGACAGAGTTCTGCGAGAAATGGAACAG 4497
Qy 485 LysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLys----- 501
Db 4498 AAGTATGAGGAACCTCAGGCTGAATTCAGGCTCCAGAGAGGATCGGTTCTCTCAGC 4557
Qy 502 -----MetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSer 518
Db 4558 ACTGAGCTGTTCAAGGTGAAGATCCCTACGAGGAATCCCTGGATCATCTTGAACACTTA 4617
Qy 519 LysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
Db 4618 AAGCGA---GAGAAAGAACTTACAAACAGGAGATTTCTGACCTACAGAGCAATTCGA 4674
Qy 539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys 558
Db 4675 GAGGTGGAAGACATATCATGAACTGGAGAAAGTAAAGAAACAACTTGTATCATGAGAAG 4734
Qy 559 AsnGluCysLeuLysSerLeuGluGlnMet-----HisGluValAlaLys 573
Db 4735 AGTGAATACAGACTTCCTTAGAGAGCAGAGGATCTTTGAGCATGAAGAGGCNA 4794
Qy 574 LysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSer 589
Db 4795 ATTCTTCGATCACTTGAG-----CTAAATCAGGTGAATCT 4833

RESULT 9
US-08-875-435B-5
; Sequence No. 5, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; PRIOR FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6644
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-5

Alignment Scores:
Pred. No.: 1,43e-05 Length: 6644
Score: 146.50 Matches: 126
Percent Similarity: 35.02% Conservative: 95
Best Local Similarity: 19.97% Mismatches: 225
Query Match: 4.56% Indels: 187

```

```

Db: 4 Gaps: 23
US-10-018-929c-3_COPY_650_1270 (1-621) x US-08-875-435B-5 (1-6644)
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2995 CAGAAGCTACAGCTCGAAAGAGTACCAGGAGCCAGAGTCAAGAAAGTTG----- 3045
Qy 42 CysThrValGluGluLysAlaLeuLeuLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 3046 -----GAGGACGACATCTCTGTCTATGACGATCAGAAACAACAAGCTCTCAAAA--- 3093
Qy 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db 3093 ----- 3093
Qy 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
Db 3094 -----GAG 3096
Qy 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
Db 3097 CGAAACCTCTCGAAGAGAGAGGATTAGTATTAAACAACAATCTTGGCCGAG----- 3147
Qy 122 GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
Db 3148 GAGGAAGAGAAGCCCAAGAACCTCACCAGCTGAAGAAACAAGCAT----- 3192
Qy 142 SerSerAspSerThrLeuPheGluAspHisIleLysLeuSerAspGluGluSerPro 161
Db 3193 -----GAATTCATGATCTCAGAACTGGAAGTGGCGCTGAGAGAGGAGGAGAGGAGC 3243
Qy 162 AsnIlePheTrpSerLysLeuGluGlyLysAsnProMetTrpLysTyrProSerAsp 181
Db 3244 CGCGAGGAGCTGGAGAAGCTG-----AAGCGAAGATGGAGCGGCGAGCCAGTGCAC 3294
Qy 182 ThrProGlnArgAsnArg-----LysArgValGlnTyrPheGluGly 195
Db 3295 CT-CCACGAGCAGATCGCCGACCTCCAGGCGCAGATCGCAGAGCTCAAGATGCAGTCGC 3353
Qy 196 SerGluAlaSerProLysThrGlyAspGlyLysAla-----Lys 209
Db 3354 CAAGAAGGAAGAGGAGCTGCAGGCGGCGCTCGCAGGCTGGAGGATGAACGCTCTCAGAA 3413
Qy 210 LysArgLysLysAlaSerAspAsp-ValThrAspProArgValThrAspProProValas 229
Db 3414 GAACAACGCCCTGAAGAAGATCCGAGGAGTGGAGGGGCACATCTCCGACCTGCAGAGGA 3473
Qy 229 p---AspAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSerProly 248
Db 3474 CCTGGACTCAGAGCGGCGCGCCAGGAAACAGAGCCGAGAAG----- 3513
Qy 248 sValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAs 268
Db 3514 -----CAGAAGCCAGA-CCTGGGGGAGGAGCTGGAGCGCTGAAGACGGA-- 3557
Qy 268 nAspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetle 288
Db 3558 -----GCTGGAGGACACCGCTGGACACCAGCCAGCAGGAGCTCCGGGCCAAGCGGG 3613
Qy 288 uAlaSerGlnAspTrpGly-LysIleProAspGluSerGlnArgArgLeuHisThrValL 308
Db 3614 AGCAGGAGGTGACGCTGCTGAAGAAGGCCCTGGAGAGAGCCCGTCCCATGTAG---- 3669
Qy 308 euLysProLysMetAlaLysLysCysGlnValLeuHisLeuSerAspAlaCysThrSerM 328
Db 3670 -----GCCCAGGTCAGGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3706
Qy 328 etValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaI 348
Db 3707 TGGTGGAGGAGCTCAGCGGAGCAGCTGGNACAGTTCNAGAGGCGCCNAGCGGACCTCGACA 3766
Qy 348 hrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuLeu----- 364

```

```
Db 3767 AGACCAAGCAGACGCTGGAGAGGAGGAGACGACAGCTGGCGGAGAGCTGGGCTCTGG 3826
Qy 365 -----VallyGlnlleuSerHislyGlnSerLeuValArgAlaAsn---SerGluL 382
Db 3827 GCCAGGCGCAAGAGGAGGTGGAGCACAAGAGAGAGCTGGAGGTGACGCTGCAGAGC 3886
Qy 382 euAlaPheLysCySeSer-----ArgValGluVal---AspTyrIleTyrS 396
Db 3887 TGCAGTCCAAAGTCAGCGATGGGGAGCGGGCGGGCGGAGCTCAACGACAAAGTCCACA 3946
Qy 396 erlleuSerCySeMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspC 416
Db 3947 AGCTGCAGATGAAGTCGAGAGCTC----- 3972
Qy 416 ysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuS 436
Db 3973 --ACGGGCATGCTCAGCGAGCGCGAGGGAAGGCCATCAAGCTGGCCCAAGGAGGTGGCGT 4030
Qy 436 ex---GlyAlaThrValArgSp----- 445
Db 4031 CCTCGGGTCCAGCTCCAGATACCAGGAGCTGCTCCAGAGAAACCGCGGAGAGC 4090
Qy 445 leAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgC 465
Db 4091 TCAACGTGTCCACCAAGCTCGCGCAGCTGGAGGAGGAGGAGAACAGCAGCTGCAGGAGCAGC 4150
Qy 465 ysSerHisThrSerThAlaThrArgAspIleGluLysThrIleSerGlyIle----- 482
Db 4151 TGGACGAGGAGATGGAGGCCAAGCAGAACCTGGAGCGCCACATCTCCACCCTGAACATCC 4210
Qy 482 ----- 482
Db 4211 AGCTCTCCGACTCAAGAAAGAGCTGCAGAGACTTTGCCAGCACCGCTGGAGTCTTGAGG 4270
Qy 483 -----LysLysLysTyTyLysGlnValGlnLysLeuValGlnGluHisGluLysL 501
Db 4271 AAGCAAGAAAGAGGTTCCAGAGGAAATTTGAGAGCTTCCAGCAGCAGTACGAGAGAA- 4329
Qy 501 ysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerV 521
Db 4330 -----GCAGCTGCTTACGATAACTGGAAAAGACCAAG--- 4362
Qy 521 alGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuL 541
Db 4363 -----AACGGCTTCAGCAGGAGCTGCAGCAGCTGG 4393
Qy 541 ysLeu---LeuAspHis-----AsnTyrGluArgLysP 551
Db 4394 TCGTAGACTTGGATAACAGCGGCACTGGTGTCCAACTGGAAAAGAGCAGAGAAGT 4453
Qy 551 heAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMethHisGluV 571
Db 4454 TCGATCATGTTGTAGCCGCGAGGAAAAGAACATCTCTTCCAAAGTATGCGGATGAAAGGACC 4513
Qy 571 alAlaLysLysLeuAlaGluAspGlu 580
Db 4514 GAGCCGAGCTGAAGCAAGGGAAGGAA 4542
```

RESULT 10

```
US-09-949-016-5562
; Sequence 5562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
```

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5562
; LENGTH: 8590
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-5562

Alignment Scores:
Pred. No.: 3,37e-05 Length: 8590
Score: 145.00 Matches: 134
Percent Similarity: 36.43% Conservative: 117
Best Local Similarity: 19.45% Mismatches: 250
Query Match: 4.51% Indels: 188
DB: Gaps: 33

US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-949-016-5562 (1-8590)

Qy 18 vallyHisValGluLyslleLyslleGlu-----SerCySeSer 30
Db 2355 CTGAAGAAATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTTACAAAGT 2414
Qy 31 GluArgThrLysllePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 2415 GAGCTTACTTAAA-----AAATCAGAGAGAGGTATTTCAG 2447
Qy 51 LeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 2448 TTACAAGAGCAGATAAATAACACAGGGTTTAGAAATCGAGAGTCTA---AAGACAGTATCC 2504
Qy 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
Db 2505 CATGAAGCTGAAGTCCATGCCGAAAGC---CTGCAGCAGAAATTTGGAA-----AGC 2552
Qy 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
Db 2553 AGCCTACTACAAATTCCTGCTGCTAGAACATCTAAGAGAAATTCGAACCT---AAACTGGAT 2609
Qy 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluAsnGluValLysLeuCysLeu 130
Db 2610 GAACTGCAAAACATCATAGCAAAAAG-----GAAAGACAGCTTAGTACTTCTTCGA 2663
Qy 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu 150
Db 2664 CAACTTAGTGAGAAAGAGAGCAGCTCTCACTAAAATACAGACAGAGATAATAGAACAAAGAA 2723
Qy 151 AspHislleLysLeu-----SerAspGluGluSerProAsn 162
Db 2724 GATTTAATTAAGGCTCTGCATACACAGCTAGAAATGCAACCAAGAGCATGATGAGAGG 2783
Qy 163 IlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAspThr 182
Db 2784 ATA-----AAGCAGCTACAGGTGGAACTTTGTGAAATGAAGCAAAACCAAGAGAGATT 2837
Qy 183 ProGlnArgAsnArgLysArgValGlnTyr-----PheGluGlySerGluAlaSer 199
Db 2838 GGAGAAAGAAAGTAGCAAGCAACAAATACAAAGGAAACTGCAAGCTGCCCTTATTTC 2897
Qy 200 ProLysThrGlyAspGlyGlyAsnAlaLysLysLysLysAlaSerAspAspVal--- 218
Db 2898 CGAAAGAACCA-----CTAAAGAAACAAAGTCTCCAAGAGGAATTGTCT 2945
Qy 219 -----ThrAspProArgValThrAspProValAsp-----AspAspGluArg 233
Db 2946 TTGCCAGAGAGTACCATTGAACGCTCACCAAGTCTCTGCGAGATGTGGAAGAGAGATT 3005
Qy 234 LysAlaSerGlyLysAspHis-----MetGlyAlaLeu-----Glu 245
Db 3006 TCTGCTCAAAATAAGAAAGAAAGATACGGCTTTAGGAGGTAGTCTTCTTCTTCAAGAGAA 3065
```

```
Qy 246 SerProlyValIleThr-----LeuGlnSer 254
Db 3066 AGAGACAACTCATACAGAAATGAGCAGGCTTTATTGGAAATACAGAGTCTCAGCAGC 3125
Qy 255 SerCysLysSer-----SerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGly 272
Db 3126 TCCTGTGAAGTCTAAACTAGCTCTAGAGGGTCTTACTGAGACAAGGAA-----3176
Qy 273 LeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSer-----290
Db 3177 -----AAGTTAGTGAAGAAATGAAATCTTTGAAATCTCTTAAGATTCGCAAGAACTACT 3230
Qy 291 -----GlnAspTyr 293
Db 3231 GAGTGGCAAGAGAAACACAAAGAGCTACAAAAGAGTATGAAATCTTCTGCAGCTCTAT 3290
Qy 294 GlyLysIleProAspGluSerGlnArgLeuHisThrVal-----307
Db 3291 GAGNATGTTAGTATGAGCAGAAAGGATTCAGCATGTGTGGAGCTGTGAGGCAAGAG 3350
Qy 307 -----307
Db 3351 AAACAAGAACTGTATGCAAGTTAAGAAGCACAGAGCAAAACAAGAGGAGACAGAAAAG 3410
Qy 308 -----LeuLysProLysMetAlaLysLeuCys 316
Db 3411 CAGTTCAGGAAGCTGAGCAAGAAATGGAGAAATGAAAGAAAGATGAGAAAGTTTGTCT 3470
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 3471 AAA-----TCTAAACAGCGAANAATCCTAGAGCTCGAA 3503
Qy 337 IleGluAsnHisArgIleTyrGluGlu-----ProAlaThrThrPheGlnAlaPheGln 354
Db 3504 GAAGAGAATCACCCTTACGGCAGAGGTGCACCTGCAGGAGAT-----3548
Qy 355 IleAlaLeuSerTrpIleAlaLeuValLysGlnIleLeuSerHisLysGluSer 374
Db 3549 -----ACAGCTAAAGAGGTGTATG-----GAAACA 3572
Qy 375 LeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyr---393
Db 3573 CTTCCTCTTCCAAATGACCATGAGAGAACTTGAAGGGTCAAAATGGAGATGATGAA 3632
Qy 394 -----IleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHis 408
Db 3633 ACCCTTCTTAAGAGTTTCAGTCTTTAATGCTCAGAAAGACTCTCTA---AGTGAAGAG 3689
Qy 409 ThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValSerThr 428
Db 3690 GTTCAAGATTTAAAGCATCATGATAGAAGGTAAATGTATCTAAACAAGCT-----3737
Qy 429 LysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn-----446
Db 3738 -----AACCTAGAGCCCGAGAAACATGATACCAACGAATGTCACT 3782
Qy 447 -----ThrLysSerMetArgAsnSerSerGluAspGluGluCysMet-----460
Db 3783 GAAGAGGGAACACAGTCTATACCAAGGTGAGACTGAGAGCAAGACTCTCTGATGATGAGC 3842
Qy 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 3843 ACAAGACTACATCTTCAGAAATCGGTTCCATCAGCAAGAGTGCACCCCTGCTGTAAAT 3902
Qy 481 GlyIleLysLysLysTyrLysLys-----GlnValGlnLysLeuValGlnGluHisGlu 498
Db 3903 -----AAGGATTTACAGCTCATCATGATGAAATTAATACTACCTACAGCAGATTGAT 3953
Qy 499 GlnLysLysMetGluLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSer 518
Db 3954 CAGCTCAAGAAAGAAATGCTGGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4013
Qy 519 LysSerValGluAlaAlaIleArgIleThrCysSerArgThrSerThrGlnValGly 538
```

```
Db 4014 CAGACTTTAGAAAATGAGAAA---AATACCTTACTGAGTCAGATATCAACAAAGATGGT 4070
Qy 539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys 558
Db 4071 GAACATAAATGCTTCAGAGGAGGTA---ACCAAAATGAACCTGTTAAATCAACGAAATC 4127
Qy 559 AsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGlu 578
Db 4128 CAAGAAGAACTCTCCAGAGTTTACCAAACTAAAGAGAGACAGCAGAAAGAGAAAGATGAT 4187
Qy 579 AspGluAlaCysTrpIleAsnArgIle 587
Db 4188 TTGGAAGAGAGCGCTTATGAATCAATTA 4214

RESULT 11
US-09-949-016-17304
; Sequence 17304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17304
; LENGTH: 38575
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(38575)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-17304

Alignment Scores:
Pred. No.: 0.00047 Length: 38575
Score: 145.00 Matches: 134
Percent Similarity: 36.43% Conservative: 117
Best Local Similarity: 19.45% Mismatches: 250
Query Match: 4.51% Indels: 188
DB: 4 Gaps: 33

US-10-018-929C-3_COPY_650_1270 (1-621) x US-09-949-016-17304 (1-38575)
Qy 18 ValLysHisValGluLysIleLysIleGlu-----SerCysSer 30
Db 4355 CTGAAGGAAATAGAGGCTGAGAAAGTAGAGTTAGAAATGAAAGTTAGTTCTTCTACACAAGT 4414
Qy 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 4415 GAGCTTACTAAA-----AAATCAGAAAGAGGATATTTTCAG 4447
Qy 51 LeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 4448 TTACAGAGCAGATTAATAAACAGGGTTTAGAAATCGAGAGTCTA---NAGACAGTATCC 4504
Qy 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
Db 4505 CATGAAGCTCAAGTCCATCCATCCGAAAGC---CTGCAGCAGAAATTTGGAA-----AGC 4552
Qy 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
Db 4553 AGCCAACTACAAATTCCTGCCCTAGAACATCTTAAGAGAAATTGCAACCT---AAACTGGAT 4609
```


; LOCATION: (1)...(119153)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-949-016-12378

Alignment Scores:				
Pred. No.:	0.0034	Length:	119153	
Score:	145.00	Matches:	134	
Percent Similarity:	36.43%	Conservative:	117	
Best Local Similarity:	19.45%	Mismatches:	250	
Query Match:	4.51%	Indels:	188	
DB:	4	Gaps:	33	

US-10-018-929C-3 COPY 650 1270 (1-621) x US-09-949-016-12378 (1-119153)

QY	18	valLyEHiSvalGluLyslelySleGlu-----SerCySer	30
DB	36933	CTGAAGCAATAGAGGCTCGAGAAAGTAGAGTTAGAAATTTAGAAAGTTAGTTCTTACACAAGT	36992
QY	31	GluaGThrLySlePheArgLeuTySerValCySThrValGluGluLySAlaLeuLe	50
DB	36993	GAGCTTACTATA-----AAATCAGAAGAGGTATTTCAG	37025
QY	51	LeuAlaArgGlnAsnMetArgGlnAsnLySAlaValGluAsnLeuAsnArgSerLeuThr	70
DB	37026	TTACAAGACGATATAATAACAGGGTTTAGAAATCGAGAGTCTA--AAGACAGTATCC	37082
QY	71	HisAlaLeuLeuMetTrpGlyAlaSerTyLeuPheAspLyLeuAspHisPheHisSer	90
DB	37083	CATGAAGCTGAAGTCCATGCCGAAGC---CTGCAGCAGAAATTCGAA-----AGC	37130
QY	91	SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis	110
DB	37131	AGCCAACTACAAATTCCTGGCTTAGACATCTAAGAGAAATTCGAACCT---AAACTGGAT	37187
QY	111	GlupheSerSerIleLeuSerSerLySgLyGlyGluGluAsnGluValLyLeuCysLeu	130
DB	37188	GAACTCAAAACTCATAGCAAAAG-----GAAGAAGAGCTTAGTACCTTTCTCGA	37241
QY	131	LeuLeuGluAlaLyHisAlaGlnGlyThrTySerSerAspSerThrLeuPheGlyGlu	150
DB	37242	CRACTTAGTGAGAAAGACGAGCTCTCACTAAATACAGACAGAGATAATAGAACAGAA	37301
QY	151	AspHisIleLySLeu-----SerAspGluGluSerProAsn	162
DB	37302	GATTTAATTAAGGCTCTGCATACACAGCTAGAAATCGAAGCAAGAGAGCATGATGAGAGG	37361
QY	163	IlePheTrpSerLySleuLeuGlyGlyLySAsnProMetTrpLySThrProSerAspThr	182
DB	37362	ATA-----AAGCAGCTCAGGTGGAACTTGTGAAATGAACAAAAACCCAGAGAGATT	37415
QY	183	ProGlnArgAsnArgLySArgValGlnTy-----PheGluGlySerGluAlaSer	199
DB	37416	GGAGAAGAAAGTAGACCAAGCAACAATACAAAGGAAACTCGAAGCTGCCTTATTTC	37475
QY	200	ProLyThrGlyAspGlyAsnAlaLySlySArgLySAlaSerAspVal---	218
DB	37476	CGAAAAAGACGA-----CTAAAAGAAAAACAAAGTCTCCAAGAGGAATTGTCT	37523
QY	219	-----ThrAspProArgValThrAspProProValAsp---AspAspGluArg	233
DB	37524	TTGGCCAGAGGTACCAATGAACGTCTCCACCAAGTCTCTGGCAGATGTGGAAGCAAGATT	37583
QY	234	LySAlaSerGlyLySAspHis-----MetGlyAlaLeu-----Glu	245
DB	37584	CTGTCTCAAAATTAAGAAAAAGATACGGTCTTAGGAAGGTTAGTCTTCTTCAAGAAAG	37643
QY	246	SerProLySValIleThr-----LeuGlnSer	254
DB	37644	AGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAAATCAGAGTCTCAGCAGC	37703
QY	255	SerCysLySer-----SerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGly	272
DB	37704	TCCTGTCAAAGTCTAAACTAGCTCTAGAGGCTTACTTACTAGACAAAGAA-----	37754


```
Db 3888 AGTCAGCGATGGGGAGCGCGCGGAGCTCAATGACAAAGTCCACAAAGCTGCAGC 3947
Qy 399 erCysMetLysSerLeu-----PheLeuGluHisThrGlnGly-----LeuG 413
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3948 ATGAAGTTGAGCGCTCACAGGATGCTTAACGAGGCGGAGGAGCCATTAAAGCTGG 4007
Qy 413 lnPheAspCysPheGlyThrAenSerLysGlnSerValValSerThrLysLeuValAenG 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4008 CCAAGGACGTGGCGCTCCCTCAGTTCCTCCAGCTCCAGGACACCCAGAGCTGCTTCAAGAAG 4067
Qy 433 luSerLeuSerGlyAlaThrValArgAspGluLysIleAenThrLysSerMetArgAens 453
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4068 AAACCCGCGAGAAGCTCAACGTGTCTACGAAGCTGCCAGCTGCAGGAGGAGCGGAACA 4127
Qy 453 erSerGluAsp-----GluGluCysMetThrGluLysArgCys-----SerHisTyrS 469
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4128 GCCTGCAAGACCAAGCTGGACGAGGAGATGGAGGCCAAGCAGAGAAGCTCGAGCGCCACATCT 4187
Qy 469 erThrAlaThr-----ArgAspIleGluLysThrI 479
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4188 CCACTCTCAACATCCAGCTCTCCAGCTCGAAGAAAGAGCTGCGAGACTTTGCCAGCACCG 4247
Qy 479 leSerGlyIle-----LysLysLysTyrLysLysGlnValGlnLysLeuValGlnG 496
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4248 TGGAGCTCTGGAAGCGGGAGAGAGAGGTTCCAGAGGAGATCGAGAAGCTCACCACGC 4307
Qy 496 luHisGluGluLysLysMetGluLeuLeuAenMetTyrAlaAspLysGlnLysLeuG 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4308 AGTACGAGGAGAAG-----GCGCGCGCTTATGATAAACTGG 4343
Qy 516 luThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrG 536
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4344 AAAGACCAAG-----AACAGGCTTCAGCAGG 4370
Qy 536 lnValGlyAspLeuLysLeu-----LeuAspHis 546
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4371 AGCTGGACGACCTGGTTGTTGATTGGACAACCGCGCAACTCGTGTCCAACCTGGAAA 4430
Qy 546 snTyrGluArgLysPheAspGlu---IleLysSerGluLysAenGluCysLeuLysSerL 565
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4431 AGAAGCAGAGAGAAATTTGATCAGTGTGTAGCCGAGGAGAGAAAACATCTCTTCCAAATACG 4490
Qy 565 euGluGlnMetHisGluVal---AlaLysLysLysLeuAlaGluAspGluAla-Cys--- 582
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4491 CGGATGAGGAGGACAGAGCTGAGCGACAGCCAGGAGAGAGAACCAAGCCCTGTGCC 4550
Qy 583 TrpIleAsnArgIleLys---SerTrpAlaAlaLysLeuLysValCysValProIleGln 601
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4551 TGGCTCGGGCCCTTGAAGAGGSCCTTGGAGCCAAAGAGAGAACTCGAGCGGACCAACAAA 4610
Qy 602 SerGlyAsnAenLys 606
Db 4611 TGCTCAAAGCCGAAA 4625
```

RESULT 15

```
US-09-949-016-1241
; Sequence 1241, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
```

```
; SEQ ID NO 1241
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1241

Alignment Scores:
Pred. No.: 6,76e-05 Length: 6861
Score: 141.00 Matches: 142
Percent Similarity: 35.68% Conservative: 96
Best Local Similarity: 21.29% Mismatches: 229
Query Match: 4.38% Indels: 202
DB: 4 Gaps: 32

US-10-018-929c-3_COPY_550_1270 (1-621) x US-09-949-016-1241 (1-6861)
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2987 CAGAAGCTGCAACTTCGAAAGGTCACGGCTGAGGCCCAAGATCAAGAAACTG----- 3037
Qy 42 CysThrValGluGluLysAlaLeuLysLeuAlaArgGlnAenMetArgGlnAenLysAla 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3038 -----GAGATGAGATCTCTGGTCATCGATGATCAGAACAAATAAATCATCAAAA--- 3085
Qy 62 ValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3085 ----- 3085
Qy 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3086 -----GAA 3088
Qy 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3089 CGAAAACTCTCTTGAGGAGGAGTATTAGTACTTAAACGACAAATCTCCAGAA----- 3139
Qy 122 GluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3140 GAGGAGAAAGGCCAAGATCTTACCAAGCTGAAAAACAGCAT----- 3184
Qy 142 SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerPro 161
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3185 -----GAATCTATGTTTTCAGAACTGGAAGTGGCTTAAAGGAAGGAAGAGAGC 3235
Qy 162 AsnIlePheTrpSerLysLeuGlyLysAsnPrometTrpLysTyrProSerAsp 181
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3236 CGACAGAGCTGGAGAAAGCTG-----AAACGGAAGCTGGAGGTGATGCCAGCGAC 3286
Qy 182 ThrProGlnArgAsnArg-----LysArgValGlnTyrPheGluGly 195
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3287 TT-CCAGAGCAGATCGTGACCTCCAGGGCCAGATCGCAGAGCTCAAGATGCGAGTGGC 3345
Qy 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSer 215
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3346 CAAGAAGGAGGAGGAGCTGCAGCGCGCTCGGCGAGCTTGACGATGAAATCGCTCAGAA 3405
Qy 216 AspAspVal-----ThrAspProArgValThrAspProVal-Asp----- 229
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3406 GAACAATGCCCTGAAGAAGATCCGGAGCTGGAGGCCACATCTCAGACCTCCAGGAGGA 3465
Qy 230 ----AspAspGluArgLysAlaSerGlyLys-----As 239
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3466 CTGGGACTAGACGGCGGCCAGGAACAAAGGCTGAAAGCAGACGAGACCTCGCGGA 3525
Qy 239 pHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSe 259
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3526 GGAGCTGGAGGCCCTAAAGACAGAGCTGGAAGACACACTGGACAGCACACCCAC- 3584
Qy 259 rGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHi 279
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3585 AGGAGCTCAGGGC-----CAAGAGGAGGAGGAGG 3611
Qy 279 sIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIlePro-AspG 299
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 .Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 12:27:00 ; Search time 983.333 Seconds
(without alignments)
3964.922 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILGSLNPSHDVKH.....SGNNKHPGSSNISONAPDV 621

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 6330945 seqs, 3139162390 residues
Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/ROBINSON929/runat_11072005_102153_29226/app.query.fasta_1.2325
-DB=Published Applications NA -QFWT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000
-USER=ROBINSON929 @CGN 1.1320 @runat_11072005_102153_29226 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	590.5	18.4	6409	19	US-10-437-963-76223 Sequence 76223, A
2	495.5	15.4	8425	19	US-10-437-963-69891 Sequence 69891, A
3	456.5	14.2	5865	19	US-10-437-963-76215 Sequence 76215, A
4	335.5	10.4	1197	20	US-10-425-115-159032 Sequence 159032, A
5	192.5	6.0	487	18	US-10-424-599-12966 Sequence 12966, A
6	166	5.2	7764	10	US-09-814-353-21548 Sequence 21548, A
7	166	5.2	7764	17	US-10-341-434-60 Sequence 60, Appl
8	166	5.2	7764	21	US-10-651-237-27 Sequence 27, Appl
9	166	5.2	7764	21	US-10-782-413-27 Sequence 27, Appl
10	161	5.0	661	20	US-10-425-115-145108 Sequence 145108, A
11	161	5.0	3045	17	US-10-369-493-26897 Sequence 26897, A
12	159.5	5.0	5883	21	US-10-956-157-859 Sequence 859, App
13	159.5	5.0	7274	21	US-10-696-909A-48 Sequence 48, Appl
14	159.5	5.0	7396	17	US-10-028-248A-35 Sequence 35, Appl
15	159.5	5.0	7396	17	US-10-107-782-35 Sequence 35, Appl
16	158.5	4.9	6354	15	US-10-084-817-158 Sequence 158, App
17	156.5	4.9	4551	18	US-10-675-072A-42 Sequence 42, Appl
18	156	4.9	5119	21	US-10-956-157-4967 Sequence 4967, Ap
19	156	4.9	5857	10	US-09-873-367C-305 Sequence 305, App
20	156	4.9	5857	21	US-10-843-641A-305 Sequence 305, App
21	156	4.9	5857	21	US-10-831-704-79 Sequence 79, Appl
22	156	4.9	5857	21	US-10-956-157-975 Sequence 975, App
23	156	4.9	5857	21	US-10-956-157-976 Sequence 976, App
24	156	4.9	6364	13	US-10-071-766-140 Sequence 140, App
25	155.5	4.8	6965	20	US-10-335-053-229 Sequence 229, App
26	155.5	4.8	7694	16	US-10-096-534-34 Sequence 34, Appl
27	155.5	4.8	7695	11	US-09-968-007A-462 Sequence 462, App
28	155.5	4.8	7695	21	US-10-843-641A-6932 Sequence 6932, App
29	155.5	4.8	8063	19	US-09-814-353-27251 Sequence 27251, A
30	152	4.7	7659	19	US-10-437-963-27251 Sequence 109076, A
31	150.5	4.7	2434	20	US-10-425-115-109076 Sequence 109076, A
32	149	4.6	5847	16	US-10-032-585-6611 Sequence 6611, Ap
33	147.5	4.6	4944	16	US-10-032-585-6572 Sequence 6572, Ap
34	147.5	4.6	5676	19	US-10-437-963-58717 Sequence 58717, A
35	147	4.6	5641	16	US-10-032-585-6646 Sequence 6646, Ap
36	146	4.5	5215	19	US-10-437-963-336 Sequence 336, App
37	144	4.5	2648	19	US-10-437-963-85842 Sequence 85842, A
38	143	4.4	4470	15	US-10-205-219-6 Sequence 6, Appli
39	141.5	4.4	83698	19	US-10-416-898-9 Sequence 9, Appli
40	141	4.4	1865	18	US-10-425-114-3516 Sequence 1602, Ap
41	141	4.4	3388	9	US-09-954-456-1602 Sequence 245, App
42	141	4.4	3388	9	US-09-967-768A-245 Sequence 988, App
43	141	4.4	3388	9	US-09-954-531-988 Sequence 1382, Ap
44	141	4.4	3388	9	US-09-954-531-1382 Sequence 85, Appl
45	141	4.4	3388	10	US-09-873-367C-85 Sequence 85, Appl

ALIGNMENTS

RESULT 1
US-10-437-963-76223
; Sequence 76223, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER FILING DATE: 2003-05-14
; CURRENT OF SEQ ID NOS: 204966
; SEQ ID NO 76223
; LENGTH: 6409
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76236C.1
; US-10-437-963-76223

Alignment Scores:
Pred. No.:      2e-50      Length:      6409
Score:          590.50    Matches:      175
Percent Similarity: 45.21% Conservative: 127
Best Local Similarity: 26.20% Mismatches: 239
Query Match:      18.36% Indels:      127
DB:              19      Gaps:       20

US-10-018-929c-3_COPY_650_1270 (1-621) x US-10-437-963-76223 (1-6409)
Qy      3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db      2263 GATGCCATCATATATTTAGTGTGATTGGAATCCAAACCAATGACTTGAGAGTTCTCCAG 2322
Qy      23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db      2323 AGGATCAGTAGAGTCTCAATCTGAATGTGTGCTATTTTTCGCTTGTATTCGCTGTGT 2382
Qy      43 ThrValGluGluLysAlaLeuLysLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db      2383 ACCGTGGAGGAAAGACTTTATACAGGAAAGCATCATATCTTTGACAGCAAGCTC 2442
Qy      63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyrLeuPhe 82
Db      2443 CAGAAATGTAATGCTATCGTGAGGCCATTCCTGCTAGTTGGGTGTCATCGTTTCTCTTC 2502
Qy      83 AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
Db      2503 AATGAGCTTGAGGAGTTCCAAAAGCATGACTACTCCAGTAAAGATTCTGAGGATGATGGT 2562
Qy      98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
Db      2563 CTC-----TTCATGAATAATGTATTTTGGATTGAGTTTGACGGCAACTATCC 2607
Qy      118 SerLysGlyGlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAla 137
Db      2608 ACCAACGTTGAAGCCAGCAGCACTAAAATGAAAAATCGAGTCATATCTCGAGCTCAGCAGAGT 2667
Qy      138 GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu--AspHisIleLysLeuSer 156
Db      2668 GGATCGGTTTTATTCTAGAGATATTGCTGTATTATAGCGAGAGGGAGGAATATCTGCAAGTT 2727
Qy      157 AspGluGluSerProAsnIle-----PheTyrSerLysLeuLeuGlyGlyLysAsnPro 174
Db      2728 GATGGTGAATGGCAAGTTCTGGACCTCTCGTCAAAATTTACTAGTGGAGGCTCTCT 2787
Qy      175 MetTyrLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
Db      2788 CATTGGCAATATATATCTGAGCCAGTCGCAAGGAACCGTAGGAAGATACAAAATATGGAA 2847
Qy      195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAla 214
Db      2848 GACCAATAGAAATACCTCGCAGAGAAACTGATGAAGCAATAATGAACCTGAGAAAATC 2907
Qy      215 SerAspAspValThrAspProArgValThrAspProValAspAspAspGluArgLys 234
Db      2908 GGAAG-----ATCATGGAT----- 2922
Qy      235 AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSer 254
Db      2923 -----TCATCTCCGAAGATCCTACCTGCTTCTGTATGTCATTTTGAAT 3891
```

```
Qy      255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
Db      2953 -----AATGATGCTGTG----- 2964
Qy      275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAsp----- 292
Db      2965 -----CTGCGCTGAAATAGTACAGCATCCAGTTCTCATGAA 3000
Qy      293 -----TrpGlyLysIleProAspGluSer-----GlnArg 302
Db      3001 ACATCAGTTGATGACACACTTGGCAAGAGCTAGGGCGAGAAAGCCCTCAGGGCACACAGAAA 3060
Qy      303 ArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSer 322
Db      3061 GGTCTTTCACACCACTCAAGCCAGAGCTGTCAAGATTGTATGATTTGTGAATTTGCCG 3120
Qy      323 AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIle 342
Db      3121 GAGACTGTCAAAATGTTTATGTGAGGAATTACTTTGATTACATTTTGAAGAATCATCAAGTC 3180
Qy      343 TyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerThrIleAlaAla 362
Db      3181 AGTCAGGAGCCAAAGGCGCATATTGCGATGCATTCACATTTGCCCTGTGTGGCGTGTGCT 3240
Qy      363 LeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeu 382
Db      3241 TCTCTTTAAAGCATTAAGATCAATCGAAGAGTCACTTGCCCTTCTGTGAGAAACTTG 3300
Qy      383 AlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLys 402
Db      3301 AACTGAGTCGCGATCAGTGCCTCGCTGAATATGTTTATGAGAAGCTAAGGATCCCTCAAG 3360
Qy      403 SerLeuPheLeuGluHisThr-----GlnGlyLeuGlnPheAspCysPhe 417
Db      3361 AAAAAGTTTTCGCGTAGAGCAAGTGAACCAAGCAAGCAGAGCCAACTTACTCCAGTAAAT 3420
Qy      418 GlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSer-----Leu 435
Db      3421 AACACATCATCTTACACACACAGACTCACCGAAATTGAGAAGTACCGGATCAATTTGT 3480
Qy      436 SerGlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGlu 455
Db      3481 CATCAGGTAACGCAATTTGATGTGTTGGAAAAATGTTTTCATCATGAAGAGGCTCCACAT 3540
Qy      456 Asp-----GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThr 472
Db      3541 GATATCTTGACCGAGGAGATGATATTAGAACAGAG----- 3576
Qy      473 ArgAspIleGluLysThrIleSerGlyIleLysLysLysTyrLysLysGlnVal----- 490
Db      3577 -----GAATTGATATCTGTTCTAGAAACTCATAGAGAAGACATGTTTAAAGG 3624
Qy      491 GlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAsp 510
Db      3625 GATGAATCTTCTGAAAGATCACAGAGAAAAAGAAATTAATTAATTAACATGTTTCTCC 3684
Qy      511 LysLysGlnLys-----LeuGluThrSer 518
Db      3685 TTAAGAGAAAAAGAAATATTCCAGGATAAACAGGAACGAGACCACACTGTTTAGCATGTCAC 3744
Qy      519 LysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
Db      3745 AACAAAAAGGAAAGTGGCAAGCTGCAAGAAACATGCAAT----- 3783
Qy      539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys 558
Db      3784 -----TTGTCGTGGAAACATCTTCGTAAGAGTGCACATTGAT-----TCAGAGGACAGG 3831
Qy      559 AsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGlu 578
Db      3832 GATGCCACCGTGAAGCTAATAATTAATGTTGCTTTTACTTTGCTTCTGTATGTCATTTTGAAT 3891
```



```
QY 579 AspGluAlaCysTrpIleAsnArgIleIys-----SerTrpAlaAlaLys 593
Db 3892 CACATGAGATGCCAGCATATCAAGTTGAAGTCAGCAATCAACTTCTATGCAACAAGGAG 3951
QY 594 LeuLysValCys--ValProIleGlnSerGlyAsnAsnLysPheSerGlySerSer 612
Db 3952 TTACAAATGAGGAATTTCTCTCAGCAACCAAAATCTGGCCACTTAGATCGTAGCTTT 4011
QY 613 AsnIleSerGlnAsnAlaProAsp 620
Db 4012 GATCAACAAATTCCTTTACCAGAT 4035

RESULT 2
US-10-437-963-69891/c
; Sequence 69891, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69891
; LENGTH: 8425
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70516C.1
US-10-437-963-69891

Alignment Scores:
Pred. No.: 2,9e-40 Length: 8425
Score: 495.50 Matches: 176
Percent Similarity: 40.69% Conservative: 108
Best Local Similarity: 25.21% Mismatches: 255
Query Match: 15.41% Indels: 159
DB: 19 Gaps: 20

US-10-018-929c-3_COPY_650_1270 (1-621) x US-10-437-963-69891 (1-8425)
QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 5785 GATGCCATCATATATGTTCTGACACCAACCCACTAAAGGCTTGAAGGCTCTTCAA 5726
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 5725 AAGATCAAGATAGAGTCACAAATTTGACCGTGTGAGCATTTTTCGCTTGTACACTCTTT 5666
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 5665 ACAGTTGAGGAGAGAGATCTTGTTCTTGGCAGGCGAGGCAATTTGTTATTGATTAACAACAT 5606
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 5605 CAAGATCTAAGGACTAGTTTAAACATTTCTTACTCAGATGGGGTGTGCAATTTCTCTTC 5546
QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
Db 5545 AGTACAGTTCAGGTCAGGTCAGCAGATGACCATCAAGTAAAGTTCTTGAATGGAGAGA 5486
QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlyGlu 122
Db 5485 CACTTTTATGATGAAGTAATCGTTGAGTTCTTTAACAAGATTATCCCAACACTGTTGAAGAC 5426
```

```
QY 123 GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
Db 5425 AGCACTGAAGTACACAGGAGTCCATAAGCAAGCACTAATATGATGGGGCACTATATTTCA 5366
QY 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
Db 5365 AGAATATTAATCTGATGGGTGAAGAGGAGGAATCTCTGTGCTTGAGGATTAATCCAGCT 5306
QY 163 IlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182
Db 5305 GAGTTCTGTTAAATTTATTGGATGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 5246
QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202
Db 5245 CTGCAATCAAGAGTTACAAAGTCCCAAGCATGGAT-----GAGGTCAATGTCGCGCT 5192
QY 203 GlyAspGlyGlyAsnAlaLysLysArgLysAlaSerAspAspVal-----Thr 219
Db 5191 GAAGAAATCAATGAAGCCAGAAAGACATAGAAAGTTGGTGAGATTATGGGTTTCATCT 5132
QY 220 AspProArgValThrAspProProValAspAsp-----AspGluArgLysAlaSer 236
Db 5131 TCAAAAGTTGTATCTGACAAAAGTAACGATGATGCTGCTGATATTTGTACTACATCT 5072
QY 237 GlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 5071 GGCCT-----GCTCTTCAA---CCAGTTGATGTCCAGCAGCAAAATCAAGTT 5027
QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
Db 5026 CAATCTGAAGGTCGGAAGCCCTGATG-----5000
QY 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
Db 5000 -----5000
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4999 -----AGTACACCCCAAGATCTTCATGCCCAAATGAAGCAAGATTGTCGAATTAAT 4946
QY 317 GlnValLeuHisLeuSer-----322
Db 4945 AAGGTGTTTACACTACGATACATACATACATGTTAATGGACGCAAAATAGCAGGCAAGTT 4886
QY 323 -----AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
Db 4885 CTCATTGAAGTAGACAATGTTACGTTGCTGCTGGTGGAGCAGTCTTCTGAATATCTTTGAAT 4826
QY 339 AsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAlaLeuSer 358
Db 4825 AATCATGTAGTTGTTTCAGGAGCCAAAGTACATATTTTCATGCGTTGAAACATAGCTTTGT 4766
QY 359 TrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAla 378
Db 4765 TGGCGAGTTCCCTCTATTCTAATACTTAAAGTGGATCATAAAGAGTCACTTGCCTTGTCT 4706
QY 379 AsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398
Db 4705 GAAAAAAGGTTAAATACGAGTGCATGAGGAGCTGCAAGGTTGGTCTATGACAGCCTA 4646
QY 399 SerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGly 418
Db 4645 -----AAGAGAAAAATTTCCAAAGAAAGACAGGTCGAACCTGGCAGCAATTTGCAATCA 4595
QY 419 ThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGlnSerLeuSerGlyAla 438
Db 4594 ACA-----TCAGTAGAAAAAGACCAACCTTTCAGCAGGAGGAGCCTCTAAT--- 4550
QY 439 ThrValArgAspGluLysIleAsnThrLys-----448
Db 4549 ATCTTGAAATCAACCACTTTTCCCTAACGAGAGATGGATCTTTCATGATATTTTCATG 4490
QY 449 -----SerMetArgAsnSerSerGluAspGluGlu 458
```

```
Db 4489 AATGGTGCACCTTCAAGAGCGCTCATTTGTTGCTGCACAAATGCTTTCAGAAGACGAGAA 4430
Qy 449 CyMet-----ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 4429 CTGATAGCTGTTCCAGGAACATCATGGAATGC---CATTTTCAACTGATGAACCTCCT 4373
Qy 475 -----IleGluLysThrIleSerGlyIleLysLysLysTyrLysLysGlnValGlnLys 492
Db 4372 GACATTTGTGGAGAAAGAAATCAACTTAATTGACAAATGTTTCTCTGTAAGGAATACCGT 4313
Qy 493 LeuValGlnGlnHisGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLys 512
Db 4312 ATCTTTGATAAACAACAGCTCTCAGATTTCCAGATTGAGAGAAATACACACAAATAAAACT 4253
Qy 513 GlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArg 532
Db 4252 GCCAGACTGAAACA-----GTGTCAAT--- 4229
Qy 533 ThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAsp 552
Db 4228 -----TTAGTTCTGGAACATATTTGTGAGAGTCACTGCTGAT 4193
Qy 553 GluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisGluValAla 572
Db 4192 -----GTAGAGACCCGAATGACACAAATAAAACAAACTGTTCACTGCTGCTG 4139
Qy 573 LysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSer----- 589
Db 4138 ATGTATGCAATTTCTGGAGCACAATGAGACTCCAGACAGCAAACTTGAGTCACCTGCAATCT 4079
Qy 590 -----TrpAla-----AlaLysLeuLysValCysValProIleGlnSerGly 603
Db 4078 AATACATGGGCTGAAGAGCGACAGCTGAGGAAAACTTTGTTGGAGCAAAATCAGGC 4019
Qy 603 ----- 603
Db 4018 CAATTAGATCACACCTTTGATCAGCAGATTGCTCTACCAGATTCAAAATTTGTTATGCAA 3959
Qy 604 -----AsnAsnLysHisPheSerGlySerSer 612
Db 3958 GAATTCATCCATTTGAAGGAACAAAGTAGCAATTCCTCATGTCAGTGGAAAGCGCT 3905
```

RESULT 3

```
US-10-437-963-76215
; Sequence 76215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76215
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1
US-10-437-963-76215

Alignment Scores: 2.09e-36 Length: 5865
Pred. No.: 456.50 Matches: 160
Score:
```

```
Percent Similarity: 41.74% Conservative: 118
Best Local Similarity: 24.02% Mismatches: 229
Query Match: 14.19% Indels: 159
DB: 19 Gaps: 24

US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-437-963-76215 (1-5865)

Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 2263 GATGCCATCATATATATTGATGATTGGAATCCAAACCAATGACTTGAGAGTTCTCCAG 2322
Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 2323 AGGATCAGTATAGAGTCTCAATCTGAATGTGTGCTCTATTTTCGCTTGTTATTCGTTGT 2382
Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 2383 ACGGTGGAGGAAAGAACTCTTACTG----- 2409
Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu--- 81
Db 2410 -----CGAAGCATG-----ATCATATTCTTGACACCAACGTCAGAACTACTT 2451
Qy 82 -----PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSer 99
Db 2452 GAGGAGTTCCAAAGCATGAC---TACTCCAGTAAAGATTTCTGAGGATGATGCTCTC--- 2505
Qy 100 PheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLys 119
Db 2506 -----TTCAATGAATATGATGATTTTGGAGTTCGACGCAAACTATCCACAAC 2553
Qy 120 GlyGluGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGly 139
Db 2554 GTTGAAGCCAGCACTAAATGGAAATGCATATCTCGAGCTCAGCAGAGTGGATCG 2613
Qy 140 ThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGlu 158
Db 2614 TTTTATTCTAGAGATATTGCTGTTATAAGCAGAGGAGGGAATATCTGCAATTGATGTT 2673
Qy 159 GluSerProAsnIle-----PheTrpSerLysLeuLeuGlyClyLysAsnProMetTrp 176
Db 2674 GATTTCCCAAAAGTTCTGGACCTTCTGTCAAATTTTACTAGTGGAGGCTCTCTCATTTGG 2733
Qy 177 LysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySer 196
Db 2734 CAATATATATCTGAGCCAGTGCAGAACCGTAGGAGATACAAATATATGGAAGCCAA 2793
Qy 197 GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAsp 216
Db 2794 ATGAGAATACCTGCAGAGAAACTGATGAAGCAATAATGAACGTAAGAAAAATCGGAGAA 2853
Qy 217 AspValThrAspProArgValThrAspProProValAspAspAspGluArgLysAlaSer 236
Db 2854 -----ATCATGGAT----- 2862
Qy 237 GlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 2863 -----TCATCTCGAAGATCTCTACCTGTCAAAGAC----- 2892
Qy 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
Db 2893 -----AATGATGCTGTG----- 2904
Qy 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAsp----- 292
Db 2905 -----CTGCCTGAATAATAGTACAGATCCAGTCTCTCATGAACATCA 2946
Qy 293 -----TrpGlyLysIleProAspGluSer-----GlnArgArgLeu 304
Db 2947 GTTGATGACACACTTGGCAAGAGCTAGGGCGCAAGAGCTTCAGGGCACACAGAAAGGCTTT 3006
Qy 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
Db 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
```

```
Db 3007 CACACCCAACTCAAGCCAGAGCTGTCAAGTTGTATGTAATTTGTTGGAATTCGCCGAGACT 3066
Qy 325 CysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGlu 344
Db 3067 GTCAAAATGTTTATGTGAGGAATTACTTGATACATTTTGAAGATCATCAA----- 3117
Qy 345 GluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuLeu 364
Db 3118 -----TGTGGCGTGTCTCTCTC 3138
Qy 365 ValIleGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPhe 384
Db 3139 TTAAGCATATAGATCAATCAAGAGAGTCACTTGCCTCTCTGTGAGAACTTGAACATAT 3198
Qy 385 LysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeu 404
Db 3199 GAGTGGCATGAAGTGTCTGCTGAATATGTTTATGAGAAGCTTAAGGATCTCAAGAAAAAG 3258
Qy 405 PheLeuGluHisThr-----GlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 3259 TTTTTCGGGTAGCAAGTGAACAAGCAAGCAGAGCCCAATCTACTCCAGTAAATAACACA 3318
Qy 420 AsnSerLysGlnSerValSerThrLysValSerValLeuValAsnGluSer-----LeuSerGly 437
Db 3319 TCATCTTACAAACCAACAGAGCTCCGGAATGAGAGTGACGGATCAATTTGTCATCAG 3378
Qy 438 AlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAsp--- 456
Db 3379 GTAACGACAATTGATGGTGAATTTGGAAAAATGTTTACATGAAGAGGCTCCACATGATC 3438
Qy 457 -----GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 3439 TTGACCGGAGGAGATCATATTAGAACAGAAAG-----GlnLys 492
Qy 475 IleGluLysThrIleSerGlyLysLysLysLysLysLysGlnVal-----GlnLys 492
Db 3469 -----GAATGTATCTGTTCTAGAACTCATAGAGAAGAACATGTTTAAAGGATGAA 3522
Qy 493 LeuValGlnGluHisGluGluLysMetGluLeuLeuAsnMetTyrAlaAspLysLys 512
Db 3523 CTTCCTGGAAGAAATCACAGAGAAAGAATTAATTAATTAACATGGTTTCTCCTTAAGA 3582
Qy 513 GlnLys-----LeuGluThrSerLysSer 520
Db 3583 GAAAGAAGATATTCAGGATAAACAAGGAAACGAGACCACACCTGTTAGACATGCACAAGCAA 3642
Qy 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540
Db 3643 AAGGAAGTGGCAAGCTGCAAGAAACATGCAT----- 3675
Qy 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
Db 3676 TTGTGTGGGAACATCTCGTAAGGTACATTCAT-----TCAGAGCAGAGGATGCC 3729
Qy 561 CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLysLeuAlaGluAspGlu 580
Db 3730 ACGGTGAAGCTAATAATTAATGAATGTTTACTTGTCTGTATGATCATTTTGAATCACATG 3789
Qy 581 AlaCysTrpIleAsnArgIleLys-----SerTrpAlaAlaLysLysLys 595
Db 3790 AGATGCCAGCATACAAGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3849
Qy 596 ValCys---ValProIleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIle 614
Db 3850 TTGAAGGAATTTCTCTCAGCAGCAAGCAAAATCTGGCCACTTAGATCTGTAGCTTTGATCAA 3909
Qy 615 SerGlnAsnAlaProAsp 620
Db 3910 CAATTCCTTTACCAGAT 3927
```

RESULT 4

US-10-425-115-159032

; Sequence 159032, Application US/10425115

```
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 159032
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1197)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_76611C.1
US-10-425-115-159032
```

```
Alignment Scores:
Pred. No.: 9.65e-25 Length: 1197
Score: 335.50 Matches: 102
Percent Similarity: 48.63% Conservative: 58
Best Local Similarity: 31.00% Mismatches: 120
Query Match: 10.43% Indels: 49
DB: 20 Gaps: 12
```

US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-425-115-159032 (1-1197)

```
Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 142 GATCCCATCATATATATTTAGTACTGGAAATCCCATGAATGATCTTGAGAGCCCTTCAG 201
Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 202 AGGTCAGCATGGAGTCACAGTCTGAACCTGTGCTATTTTCGTTTGTACTCGTCTTTT 261
Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 262 ACAGTGGAGGAAAAAGCTTATTCTTGCAGAGCATGACCATATATTTGGATAGTAATATT 321
Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 322 GTGAATATACCCCTAGCCCTGAGCCATTGTCTCAGTTGGGGCGCATCAATTTCTCTTC 381
Qy 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
Db 382 AATAGACTTGAGGAGTTGCAGCAACATAGTAC-----TCAATGTCTCTGTT 429
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 430 GATGAGCTTTTCATGGATAATAGATTGAGTTTGTGAGTTTGTGACAAAATTTACTCAGCAAGGTT 489
Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 490 GAATCAGACTGAACTCAGCAATACAGCCATACAGCCATATCATCTGTGTGGATCTTTC 549
Qy 141 TyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGluGlu 159
Db 550 TATTCAGGGCTATTGTTGAGCGGTGAGGGGAGGGAATTCATCAGTTGATGTTGATG 609
Qy 160 SerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLys 177
Db 610 CTGCCAAAAGTCTGGGCTACTGGGTAAGTTTACTTAATGGGAGGTCTCTCTCAGTGGCAG 669
Qy 178 TyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGlu 197
Db 670 TATATATCTGAGCCAGTACAAAGAACCGCTAGAAAGATA-----AATAACATGGAA 720
```

```

Qy 198 AlaSerProlyThrGlyAspGlyGlyAsnAlaIlyLysArgLysAlaSerAsp 217
Db 721 CAACAATGAAACACAGAC-----AAACAGTTGAAATAAATCTACTGAAGAA 768
Qy 218 ValThrAspProArgValThrAspProProValAspAspGluArgLysAlaSerGly 237
Db 769 ACTGATGAACCCAGAGTA-----AAGCTAGAGNAATGGT 804
Qy 238 Lys---AspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 805 GAAATCATGATTCATCTGCAATAGACTCACCT----- 837
Qy 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
Db 838 ---GGCAAAAACAGGACACATATTCGAGGAACAATACC-----CCACCT 882
Qy 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
Db 883 AGTTCTCAT-----CAATATCAGTTGAGNATACATGCGAGGAACCTA 924
Qy 297 Pro-----AspGluSerGlnArgArgLeuHisThrValLeuLysProLysMet 312
Db 925 GAGCGAAGCAATCTTTCATGCCACACAAAAAGGCTACACGTTCAACTCAAGCCAGAGATT 984
Qy 313 AlaLysLeuCysGlnValLeuHisLeu 321
Db 985 TCGAAGTTATATAAGTTGCTGCAACTG 1011

```

RESULT 5

```

US-10-424-599-12966
; Sequence 12966, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 12966
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1
US-10-424-599-12966

```

```

Alignment Scores:
Pred. No.: 2.6e-10 Length: 487
Score: 192.50 Matches: 57
Percent Similarity: 45.64% Conservative: 32
Best Local Similarity: 29.23% Mismatches: 71
Query Match: 5.99% Indels: 35
DB: 18 Gaps: 4

```

US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-424-599-12966 (1-487)

```

Qy 337 IleGluAsnHisArgIleTyrGluGluProIleThrPheGlnAlaPheGlnIleAla 356
Db 1 ATGAACAATCTTTATGTGTAATAGGGAACCGTTTTCATATATCACAGCGCTTTTCATATATCT 60
Qy 357 LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 61 CTGTGTGGACTGCAGCTCTTCTGCTGAACACACAACTTGACCCCATCTGCTCTCTATA 120
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSer 396
Db 121 CAA-----GATTTGAACTTTGAGTGTGAAGAAAGAGGTGGACTACATTTGTTCT 171

```

```

Qy 397 IleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCys 416
Db 172 ATGTGTGTGTTGTAAGAAATATCTTATATCGCACAGGAATATCATGATACCGGC 231
Qy 417 PheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSer 436
Db 232 TTT----- 234
Qy 437 GlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAsp 456
Db 235 -----ACAAAAGCATCTCGACCATCAAAACAGGGCA 264
Qy 457 GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGlu 476
Db 265 TATTCCTGTACAGAGTTGCACGA-----GAGTTGAATTTGTTCAAAAAGATATGTCC 318
Qy 477 LysThrIleSerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGlu 496
Db 319 AAAAGTATTAAAGAAATTCAGAAAGAGTGTGAAAGAGCTGAAGAAGCTACATATTTTG 378
Qy 497 HisGluGluLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGlu 516
Db 379 CAAGAAGAAGAGAGCAAGCAAGGTTGAGAGCGAGCTATTGAGGAAGAAAGGCTAATTTTGAT 438
Qy 517 ThrSerLysSerValGluAlaAlaValIleArgIleThrCysSer 531
Db 439 GAAAGGTACACATAGAGTCTGTATATTCGA---TCCTGCTCT 480

```

RESULT 6

```

US-09-814-353-21548
; Sequence 21548, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21548
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 7755, 7756, 7757, 7758, 7759, 7760,
; LOCATION: 7761, 7762, 7763, 7764
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21548

```

```

Alignment Scores:
Pred. No.: 1.07e-05 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251

```



```

Db      5208  CACACCGGAACTGCTGCTGAAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
Qy      564  -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db      5268  TCAGATCAGGAGCACATCAAACTCTGGAGGAGGAACCTATGGATTTTGTTCCAAGTTT 5327
Qy      565  -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db      5328  ACAGAAATTGATGCTCGAAACTGCATAAGTTATACAAAGATGGCTCATAGAAAAGGTCT 5387
Qy      578  GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db      5388  CAAGAGAAGAG-----GAGCNAAGAGAAAGAGACGAC 5420
Qy      598  ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db      5421  GTG-----ACTGGGGGTAAAGAAACCATTTTGTCTCAGAGGGCTCAGGCTCAGGCCG 5471
Qy      614  -----IleSerGlnAsn 617
Db      5472  GACTCTCTGATATCTCAGTCC 5492

RESULT 7
US-10-341-434-60
; Sequence 60, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (708)..(5924)
; OTHER INFORMATION:
US-10-341-434-60

Alignment Scores:
Pred. No.: 1.07e-05 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 17 Gaps: 37

US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-341-434-60 (1-7764)
Qy      2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
Db      3342  CGCGACACAGTCGTCATCTTGACTCTGAGTGGAAACCCCGACGATGACTTGCAGGCACAA 3401
Qy      22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db      3402  GCCCGCGCATAGAAATTGGTCAGAAGAAGCAGGCTAAATATTATACCGCTTAGTTACAAAG 3461
Qy      42 CysThrValGluGluLysAla-----Leu 49
Db      3462  GGGACTCTGGAGGAGGAGATCATAGACGGGCCNAAGAGATGGTATTAGATCATCTG 3521
Qy      50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64

```

```

Db      3522  GTGATTTCAGCGCATGCACACCACTGCGCGGACGATCCTCGAAAAACAACACTCAGGAAGTCC 3581
Qy      65  -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db      3582  AACTCAAAATCTTTTAAATAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGGAT 3641
Qy      81  LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db      3642  CTCTTCAAGAACTGGAA-----GGGGAGGAATCA 3671
Qy      101  GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db      3672  GAACCTCAGGAAATGGAT-----ATAGATGAAATTTTTCGCTTGGCTGAA 3716
Qy      121  GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db      3717  ACGAGAGAGAATAAGGTGTCAACAAGTGCACAGACTCACTCTTA---TCACAGTTTAAAG 3773
Qy      141  TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db      3774  GTTGCCCAACTTTCACAACATGGGAAGATGAAGACGCTA-----GAAGAGCGT 3821
Qy      161  ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db      3822  CCTCACAAAGGACTGGGATGAGATCATT-----3848
Qy      181  AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
Db      3849  -----CCAGAGGAACAAGAAAAAGTAGAGGAGGAGCGGCAGAGAGGAGCTAGAA 3902
Qy      195  -----GlySerGluAlaSerProLysThrGly 203
Db      3903  GAAATTTATATGCTGCTCGAATTCGAGGTTCCACTAAAAGGCTCAGACAAATGACAGT 3962
Qy      204  AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db      3963  GACTCTGACACTGAGTCTAAGAGCGCAGGCCAGAGATCCTCTGCTTCTGAGAGTGAACG 4022
Qy      224  ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db      4023  GAAGAC-----TCTGATGATGACAAAG-----4046
Qy      244  LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db      4047  -----CCAAAGCGCAGAGCGGCTCCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 4097
Qy      264  ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db      4098  TTTACTGATGCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT-----4145
Qy      279  HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db      4146  -----GGTCTCCCTCTTGAACCGCTGGAGTGTTCAGCATGTAT---GCTGAGCTG 4193
Qy      297  ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db      4194  GTAGATAAGTCCGTGGCAGATCTG-----AAGCGCTGGGT 4229
Qy      317  GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db      4230  GAACTGATCCAC-----AACAGCTGTGTCTCAGCAATGCGAATATGAAGAGCAGCTG 4283
Qy      337  IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
Db      4284  AAAGAAATCCCGAGCGAGGGAAGGACCGGGAAGG-----AGAGTCCCAACATCAAG 4340
Qy      357  LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db      4341  ATATCCGAGTT---CAGGTTAATGTGAATCCATTATCCACATGAAGAGGAGTTT---4394
Qy      377  ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db      4395  -----GAGATGCTGCATAAATCTATCCCTGTGCAACCTGGAAGAAAAA 4442

```

```
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 4443 TACTGCTTAACCTGCTCGTGTCAAGAGCTCACATTTTGATGTAGAGTGGGGGTGGAAGAT 4502
Qy 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
Db 4503 GATTCTCGCTGTTGCTGGGGATTTATGAACATGGCTATGGAAACTGGGAGTTAATTA 4562
Qy 422 sGlnSerValSerThrLysLeuValAsn----- 432
Db 4563 ACAG-----ACCCAGAGCTTAATAATTACTGACAAAATCTCCCGTGGAGACA 4610
Qy 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn-ThrLysSerMet 450
Db 4611 GATAAAGAGCTCAGGGAGAGCAGCTACAGCCGAGCGGATTACTTGTGTAAGCTGCTC 4670
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4671 AGAAGGGTCTGGAGAGAGAGGGGGCTGTGACAGGTGGGGAGAGGCCAAATTAAGAG 4730
Qy 464 Arg----- 464
Db 4731 CGGAAGCTCGGTAAAGAGAGAAAACAAAGTGCCAGGCTGAAGAGGACCATGGAATT 4790
Qy 465 ----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4791 GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGAGAGAGTGAAGATGAT 4850
Qy 481 GlyIleLysLysLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4851 GGCTTGAAGAAAGCTCCAAATGAAGAAAGAAAGAGAG-----AAGAAAGAGAGCAAGGAGAAC 4907
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4908 AAGGAGAAACAAATGAGTCTTAGAAGAGACAAAGAGGGGACAGAGAAAGAAAGAGTCA 4967
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4968 AAGATAAAGAAAGAGAGCCCTAAAGTGGTGATGCCAAATCTTCGAGTAAATCAAAGCGA 5027
Qy 524 ----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 5028 TCTCAGGGTCTGTCCATATTACAGCAGGAAGTGAACCTGTGCCCATTTGGAGAGGATGAG 5087
Qy 539 ----- 539
Db 5088 GATGATGATCTGGACCAAGAGACATTCAGCATATGTAAAGGAGGATGAGCCCGTGAAA 5147
Qy 540 ----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5148 AAGGCACTGAAACAGCTCGACAAACCTGACAAGGGGCTCAAGCTGCAAGAGACAGCTGGA 5207
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5208 CACACCCGGAACCTGCTGCTGAAATTCGGAGACCGGATAGCCGAGTGCCTTTAAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGGGAACCTATGGATTTTGTTCAGTTT 5327
Qy 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 5328 ACAGAAATTTGATGCTCGAAACCTCATAAAGTTATACAAAGATGGCTCATAGAAAGGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 5388 CAAGAAAGAGAG-----GAGCAAAAGAGAAAGAGAGCAGC 5420
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGTGAAGAAACCATTTTCGTCCAGAGGGCTCAGGCTCCAGCGG 5471
```

```
Qy 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 8
US-10-651-237-27
; Sequence 27, Application US/10651237
; Publication No. US20050048494A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: ABS-5003 US NP
; CURRENT APPLICATION NUMBER: US/10/651.237
; CURRENT FILING DATE: 2003-08-27
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
; US-10-651-237-27
Alignment Scores:
Pred. No.: 1.07e-05 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 21 Gaps: 37
US-10-018-929c-3_COPY_650_1270 (1-621) x US-10-651-237-27 (1-7764)
Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
Db 3342 GCGGACACAGCTGTCATCTTTGACTCTGACTGGNACCCCGAGATGATCTTCGACGACAA 3401
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3402 GCCCGAGCGCATAGATTGGTCAGAGAGAGCAGGATAATATTACCGCTTAGTTACAAAG 3461
Qy 42 CysThrValGluGluLysAla-----Leu 49
Db 3462 GGGACTCTGGAGGAGGAGATCATAGACGGGCCAAAAGAGATGGTATTAGATCATCTG 3521
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn----- 64
Db 3522 GTGATTCAGCGCATGGGACACCATCTGGCGGACGATCTCGAANAACAACTCAGGAAGGTCC 3581
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3582 AACTCAAAATCCTTTTATAAAGAGAGAGCTGACAGCTATTTTGAAATTTGGAGCAGAGAT 3641
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3642 CTCTTCAAGAACATGGAA-----GGGAGGAATCA 3671
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3672 GAACCTCAGGAATGGAT-----ATAGATGAAATTTTCGGTTCGCTGAA 3716
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3717 ACAGAGAGAGAAATGAAGTGTCAACAAGTGCAACAGATGAACCTCTA---TCACAGTTTAAAG 3773
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluLysSer 160
Db 3774 GTTGCCCAACTTGCACAAATGAAGATGAAGAGAGCTA-----GAAGAGCGT 3821
Qy 161 ProAsnIlePheTrpSerLysLeuGlyLysAsnProMetTrpLysTyrProSer 180
Db 3822 CCTCACAAGGACTGGGATGAGTCAATT----- 3848
```

QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu----- 194
Db 3849 -----CCAGAGAACAAAGAGAAAGTAGAGAGGAGCGCGCAAGAGGAGCTAGAA 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3903 GAAATTTATATGCTGCTCGAATTCGGAGTTCCTCACTAAAGAGCTCAGACAAATGACAGT 3962
QY 204 AspGlyGlyAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db 3963 GACTCTGACACTGAGTCTAAGAGCGAGCGCCAGAGATCCTCTCTCTGAGAGTGAACG 4022
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 4023 GAAGAC-----TCTGATGATGACAAAG----- 4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 4047 -----CCNAAGCCAGAGCGGCTCCGAGGAGTGTGCGGAAGGACCTCGTGGAGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 4098 TTTACTGATCAGAGATCCGAGGTTTCATCAAGGCTTATAGAGATTT----- 4145
QY 279 HisLeuSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db 4146 -----GGTCTCCCTCTGAAACGGCTGGAGTGTTCAGCAGTGAT-----GCTGAGCTG 4193
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4194 GTAGATAAGTCGTGGCAGATCTG-----AAGCGCTGGGT 4229
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 4230 GAACTGATCCAC-----AACAGCTGTGTGTCGAATGAGCAATGAGGAATATGAAGAGCAGCTG 4283
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 4284 AAGAAATATCCAGCGAGGAGAAAGACCGAGGAAAGG-----AGAGTCCACACATCNAAG 4340
QY 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 4341 ATATCCGAGTT--CAGGTTAATGTGAATCCATTATCCACATGAAGAGGAGTTT-- 4394
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db 4395 -----GAGATGCTGCTAATAATCTATCCCTGTGACCCCTGAAGAAAAA 4442
QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 4443 TACTGCTTAACCTGCTGTCGTCGAAGCTGCACATTTTGATGTAGTGGGGGGTGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer-----Ly 422
Db 4503 GATTCTCGCTGTGCTGGGGATTATGAACATGGCTATGMAAAGCTGGGAGTTAATTA 4562
QY 422 sGlnSerValValSerThrLysLeuValAsn----- 432
Db 4563 ACAG-----ACCAGAGCTTTAAATTTAACTGACAAAATTTCTCGCGTGGAGACA 4610
QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
Db 4611 GATAAAAGCCTCAGGGGAAGCAGCTACAGACCCGAGCGGATTAATTGTTGTAAGCTGCTC 4670
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4671 AGAAAGGGTCTGAGAGAGAGGGGGCTGTGACAGGTGGGGAGGAGGCGCAAAATTAAGAAG 4730
QY 464 Arg----- 464
Db 4731 CGGAAGCCTCGGGTAAAGAGGAAACAAAGTGCCAGGCTGAAAGAGGAGCATGGAAAT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480

Db 4791 GAGCTTTTCATCTCCTAGGCATTTCAGATAATCCATCAGAAGAGGAGGAGTGAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4851 GGCTTGGAAAAAGTCCAATGAAAAAACAACAGAG--AAGAAAGAGAAACAAGGAGAAC 4907
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4908 AAGGAGAACAAATGAGTTCTAGGAAGACAAAGAAAGGGGACACAGAAAGAAAGAAAGTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4968 AAGATAAGAAAGAGAGAGCTTAAAGTGGTGATGCCAAATCTTCGAGTAAATCAAAGCGA 5027
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 5028 TCTCAGGGTCTGTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGAGGATGAG 5087
QY 539 ----- 539
Db 5088 GATGATGATCTGGACACAGGAGACATTCAGCATATGTAAAGAGAGGATGAGCCCGTGAAA 5147
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5148 AAGGCACCTGAACAGCTCGACAAACCTGACAGGGGCTCAACGTCGAAGAACACAGCTGGA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer----- 564
Db 5208 CACACCCGGAACCTGCTGTGTAATAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
QY 564 ----- 564
Db 5268 TCAGATCAGAGAGACATCAAACTCTCGAGAGAGAACCTATGATGATTTTGTTCACAGTTT 5327
QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 5328 ACAGAATTTGATGCTCGAAAACTGCATAGTTATACAGATGCTCATAGAAAAAGGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLysLysValCys 597
Db 5388 CAAGAAGAAGAG-----GAGCAAAAGAAAGAAAGACGAC 5420
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGTAAGAAACCAATTCGTCCAGAGCGCTCAGGCTCCAGCCGG 5471
QY 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 9
US-10-782-413-27
; Sequence 27, Application US/10782413
; Publication No. US20050048526A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: VDX-5002 CIP
; CURRENT APPLICATION NUMBER: US/10/782,413
; CURRENT FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: 10/651,237
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
; US-10-782-413-27
Alignment Scores:


```
Db 5148 AAGGCACTGAACAGCTCGACAAACCTGACAAAGGGCTCAACGTCGAAGAACAGCTGGAA 5207
Qy 553 -----GluIleYsSerGluYsLysGlnGluYsLeuYsSer--- 564
Db 5208 CACACCGGAACTCGCTGCTGAAATCGGAGACCGGATAGCCGAGTGCCTTTAAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGGAGCAGACATCAAACTCTGGAGGAGGAACTATGGATTTTGGTTTCCAAGTTT 5327
Qy 565 -----LeuGluGlnMetHisGluValAlaLysLysLeuAla 577
Db 5328 ACAGAAATTGATGCTCGAABACTGCATAAGTTATACAAGATGGCTCATAGAAAGGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleYsSerTrpAlaLysLeuYsValCys 597
Db 5388 CAAGAAGAAGAG-----GAGCAAAAGAAAGAAAGACGAC 5420
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGGTAGAAACCAATTCGTCCAGAGGCCTCAGGCTCAGGCCGG 5471
Qy 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 10
US-10-425-115-145108
; Sequence 145108, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 145108
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1
US-10-425-115-145108
Alignment Scores:
Pred. No.: 8.44e-07 Length: 661
Score: 161.00 Matches: 50
Percent Similarity: 44.51% Conservative: 23
Best Local Similarity: 30.49% Mismatches: 75
Query Match: 5.01% Indels: 16
DB: 20 Gaps: 5
US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-425-115-145108 (1-661)
Qy 239 AspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSer 258
Db 6 GATGAAGTTGTAATCCAGCAATCACTGCTCCACAAACAAAAGAAAGTGTAAAGAA 65
Qy 259 SerGly-----ThrAspGlyThrLeuAspGlyAsnAspAlaPhe 271
Db 66 ATTCAATTCAGAAACTTCTGAGATGCCAGATGGGTGTCATCTGATTTTGATAATCAT 125
Qy 272 GlyLeuTySerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGln 291
Db 126 CATATGTGTTCTGTAAATATCAT-----GACCTGCTTCTGAAATGAGTACTCCAAAGTGCT 182
Qy 292 AspTrpGlyLysIleProAsp-----GluSerGlnArg-----ArgLeu 304
```

```
Db 183 GACCTTTCATCTACTCTGAGACAGAGTGTGAAACATGAGGACACCTAAAAGTAGTTTT 242
Qy 305 HisThrValLeuYsProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
Db 243 CACGCTGAACTTAAAGCGTGTGCTCAAGCTGCTCAAGCTGTTAAAGCTACCCAGATAAT 302
Qy 325 CysThrSerMetValGlyAsnPheLeuGluTyTrValIleGluAsnHisArgIleTyGlu 344
Db 303 GTACTTTTCTGGCCCAACAGTCCCTCGAATATCTTTAAATATCATTTAGTTGTGCGG 362
Qy 345 GluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeu 364
Db 363 GAGCCACGAGGAGCATCTGCGATGCTTCAACATAGCTGTGTGTGGGTGCTGCTCTTTT 422
Qy 365 ValLysGlnIle---LeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAla 383
Db 423 CTGAATACATCAGTGGATTCAGAAATCACTTGCCCTTCAGATGGATTGAAT 482
Qy 384 PheLysCysSer 387
Db 483 TATGAATGCAAT 494
RESULT 11
US-10-369-493-26897
; Sequence 26897, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 26897
; LENGTH: 3045
; TYPE: DNA
; ORGANISM: Neurospora crassa
; US-10-369-493-26897
Alignment Scores:
Pred. No.: 8.61e-06 Length: 3045
Score: 161.00 Matches: 131
Percent Similarity: 37.00% Conservative: 111
Best Local Similarity: 20.03% Mismatches: 237
Query Match: 5.01% Indels: 176
DB: 17 Gaps: 29
US-10-018-929C-3_COPY_650_1270 (1-621) x US-10-369-493-26897 (1-3045)
Qy 17 AspValYsHisValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePhe 36
Db 673 GACATCGACATTTGGCGGCAAGTGTCAACACCGATCTCAAAAGAGAGGAGGATTCCTATCT 732
Qy 37 ArgLeuTySerValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMet 56
Db 733 CGCTTG-----CAGGAAGAGCT-----ACTCGCAAAACCGAG 765
Qy 57 ArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrp 76
Db 766 CAGCCGACCGAGATGTCCGAATCTCTAAACCGAACGCTGAAC----- 807
Qy 77 GlyAlaSerTyrlleuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSer 96
Db 808 -----GATATCATGGATACCCGGCAGCGCTTGGCGTTCTTATCAAA 849
```

Qy 97 GlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeu 116
Db 850 GCGCGCCGCCACCTCATGATAGCGCAACAATCTTGAAGTCTTCGATACATATTG 909
Qy 117 SerSerIleGlyGlyGluGluAsnGluValIleHisGluLeuLeuGluAlaLys--- 135
Db 910 AACCTGTATC-----GAAGACACCGACCAAGCATTTGTTCAACCGAAGAACCCAGGAC 963
Qy 136 HisAlaGlnGlyThrTySerSerAspSerThrLeuPheGlyGluAspHisIle---Lys 154
Db 964 AAGGACAGAACGCGCGCAAGGAATCCGCCACACGATCCCGCAGTGAACAATCCAGAAG 1023
Qy 155 LeuSerAspGluGluSer-----ProAsn 162
Db 1024 CTCACGCTGAAACAAACAAGCTTAAAGAACAGAGATGAGCAAAAAGAAAGCCGAC 1083
Qy 163 IlePheTrpSerIleLeuGlyGlyLysAsnProMetTrpLysTyProSerAspThr 182
Db 1084 CGCTTTTACCAAACTC-----AAGAAAAGGATGGCTTGAGCCGACCGAGAT 1134
Qy 183 Pro-----GlnArgAsnArgLysArgValGlnTyrPheGluGly 195
Db 1135 CCATCGCGCTCACTTGCCTCGACAGAGTCGTGAAGACCAACAAGATCGCTGGAG 1194
Qy 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArg---LysLysAla 214
Db 1195 TCTGAAAGAAATCCCTGAAACCAAGAAATCACTAATCTCAACAATCGATCGAAGAGCC 1254
Qy 215 SerAspAspValThrAspProArgVal-----ThrAspProProValAspAsp 230
Db 1255 AATGTGCGCTCAAGGACATCAAGCTTATGACGTGGAGGAGATTTCTGTGCTGATCTA 1314
Qy 231 AspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----ProLys 248
Db 1315 CAACAACGTCAGGAA-----GGGATATCAAAAGCTTGAACCAACAG 1356
Qy 249 ValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsn 268
Db 1357 GTTCTGGACCTCAAGCAAAAATTTGGCAGCAGGAGCCAGCGCGCGCGCA-----TCT 1409
Qy 269 AspAlaPheGlyLeuTySerMetGlySerHisIleSerGlyIleProGluAspMetLeu 288
Db 1410 GAAGGATATTCAGTCTTTCGATTTGGAGAACAGTCTTTGGAGACCAAGCAGCAACGTGT 1469
Qy 289 AlaSerGlnAspTrp-----GlyLys 295
Db 1470 GCAAGCGCAAAAGAGGTCTTGACAGCAGCAGCTTTCTCAGACGAAGCAAGCTGGAAA 1529
Qy 296 IleProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetAlaLysLeu 315
Db 1530 GGTGGAGACGACCTCAAGAA-----CACGGCTCGCAAAAGATGGATT----- 1574
Qy 316 CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyr 335
Db 1575 -----GGAGACTCAAGAAATGA 1592
Qy 336 ValIleGluAsn-HisArgIleTyGluGluProAlaThrThrPheGlnAlaPheGlnI 355
Db 1593 GTGTCCAGCAAGAAAGTCTCGAGGAGAG----- 1626
Qy 355 eAlaLeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLe 375
Db 1627 -----ATCAGCAGGCTCAAGAGAGGAGGTTCAGCTTGGCCAGGAAGATTCGCG 1673
Qy 375 u-----ValArgAlaAsnSerGluLeuAlaPheLysCysSerArgVa 389
Db 1674 TGCCGCCAACATCGTCGATGATCAAAACCGAACCACCAATGTCGGAAGCATCATCTCGGT 1733
Qy 389 lGluVal-----AspTyrIleTy 395
Db 1734 CGAACAACTCATGAGGAGCAAAACCTCACTACCGTGCACCGCAGCGACTTCGTCCA 1793
Qy 395 rSerIleLeuSerCysMetLysSerLeu-----PheLeuGluHisThrGlnI 411

Db 1794 AGGACTGGTTCGCAAGTTGAAGGACGCGCAGGACACAAAGACTTGGATGCCAGTAGGA 1853
Qy 411 yLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuVa 431
Db 1854 TGTCAATTCGTGAATGCATCGC-----CTTCT 1880
Qy 431 lAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn----- 446
Db 1881 CAATCGTGTCTTGAAG---GAGCCTGTCTTTGATGAGAAGCAAGGAGATGCTTCAGAC 1937
Qy 447 -----ThrLysSerMetArgAsnSerSerGluAspGluGlu-----Cy 459
Db 1938 TTTGCAAAATGAGTTGAATGAGGCTCTGAGAAAGGCAAGGAGCGCCGCAACAAAGCCG 1997
Qy 459 sMetThrGluLysArgCysSerHisTySerThrAlaThrArgAspIleGluLysThrI 479
Db 1998 GCAGTTTCAGAGCATCATTCACAAATACGAAGCGCAAGAAAGAAAGCCCTTGAGAGGAGGT 2057
Qy 479 eSerGlyIleLysLysLysTyLysLysGlnValGlnLysLeuValGlnGluHisGluGl 499
Db 2058 C-----AAGGTTTCATAAGACGAGATCAAGAGTTGAAGCAGGAGATCACTGA 2105
Qy 499 uLysLysMetGluLeuLeuAsn-----MetTyAl 509
Db 2106 GAAGAAACCTCACTGGCAACAAAGCAGCAGAGAGAGATATGCTCAAGGAAAGCTACGA 2165
Qy 509 aAspLysGlyGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleTh 529
Db 2166 AGACCAATCAAAATCTCAATCGGACCAACAAAGCGTCTCGCGAGCTGAAAGTCAA 2225
Qy 529 rCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuAspHisAs 546
Db 2226 ACACCAAAAGCAGCTGACGAGCTCAGAAAGGACGCGACTTGAAA-----GAAACCAA 2279
Qy 546 nTyrgluArgLysPheAspGluLysSerGluLysAsnGluCysLeuLysSerLeuGl 566
Db 2280 CCTCTCGAAGTTAGACACACTCAGGACGAGAGATGAAGAGTGAAGAACAGGCTTCA 2339
Qy 566 uGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
Db 2340 AGCGGATTAGCGCTCAGAAAGCGGAGCTCAGCAAGGAT 2379

RESULT 12
US-10-956-157-859
; Sequence 859, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 859
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-859

Alignment Scores:
Pred. No.: 3,37e-05 Length: 5883
Score: 159.50 Matches: 137
Percent Similarity: 35.14% Conservative: 109
Best Local Similarity: 19.57% Mismatches: 225
Query Match: 4.96% Indels: 231
DB: 21 Gaps: 28

US-10-018-929c-3_COPY_650_1270 (1-621) x US-10-956-157-859 (1-5883)


```
Db 4245 GACGGCTGCAGCAGCTGGACGACCTGCTGGTGGACCTGGACACCCAGCCAGAG 4304
QY 507 -----
Db 4305 CGCGTGCACCTGGAGAGAGCAGAGAGATTGTGACCAAGCTCTGCGCGGAGGAGAGAC 4364
QY 508 -----TyrAlaAspLysGlnLysLeuGluThr-----SerLysSe 520
Db 4365 CATCTCTGCCAAGTATGACGAGGAGCGGCGGCTGAGCGGAGCGCGGAGGAGAGGA 4424
QY 520 rValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLe 540
Db 4425 GACCAAGGCTCTGCTGCTGGCGCGGCTGGAGGAGCCATGCGAGCAGAGGCGGAGCT 4484
QY 540 uLysLeuLeuAspHisSerThrGluArgLysPheAspGluIleLysSerGluLysAsnG1 560
Db 4485 GGAGCGGCTCAACAAGCAGTTCCGACGAGATGGAGGACCTTATGAGCTTCAAGGATGA 4544
QY 560 uCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspG1 580
Db 4545 TGTGGGCAAGT-----GTCCACGAGCTGGAGAGTCCAAAGCGGCGCTTAGAGCA 4595
QY 580 uAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro 599
Db 4596 GCAGG--TGGAGGAGATGAAGACGACGCTGGAAGAGCTGGAGGACGAGCTGCAGGCCA 4651
```

RESULT 14

```
US-10-028-248A-35
; Sequence 35, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimketa, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangoli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
```

```
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-028-248A-35

Alignment Scores:
Pred. No.: 4,77e-05 Length: 7396
Score: 159.50 Matches: 139
Percent Similarity: 35.73% Conservative: 109
Best Local Similarity: 20.03% Mismatches: 228
Query Match: 4.96% Indels: 220
DB: 17 Gaps: 29

US-10-018-929c-3_COPY_650_1270 (1-621) x US-10-028-248A-35 (1-7396)
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3017 CAGAAGCTGCAGCTGGAGAGGTGACCCAGCGGCGAAGCTGAAAAGCTG----- 3067
QY 42 CysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 3068 -----GAGGAGGAGCAGATCATCTCTGGAGGACGAGAACTGCAAGCTGGCCAAAG--- 3115
QY 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db 3115 ----- 3115
QY 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
Db 3116 -----GAA 3118
QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
Db 3119 AAGAAACTGCTGGAGACAGAAATAGCTGAGTTCCACCACCACTCAGAA----- 3169
QY 122 GluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
Db 3170 GAGGAGGAGAAATCTAAGAGCCTCGCCACAGCTCAAGACAGCAT----- 3214
QY 142 SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGlu-SerPr 161
Db 3215 -----GAGGCAATGATCACTGCTTGAAGAGCGCTCCGCGAGGAGGAGAGCAG 3265
QY 161 oAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAs 181
Db 3266 CGACAGGAGCTGGAGAGACCCCGCGGAGC-----TGGAGG-----GA 3304
QY 181 pThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLy 201
Db 3305 GACTCCACAGACTCAGCGACACAG-----TCGCCGAGCTCCAGGCCAG 3349
QY 201 sThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspPr 221
Db 3350 ATCGCGAGCTCAAGATGCAGCTGGCCCAAGAAAGAG----- 3386
QY 221 oArgValThrAspProValAspAspAspGluArgLys----- 234
Db 3387 -AGGAGCTCCAGCGCGCTTGGCCAGAGTGGAGAGAGAGAGTCCCAAGAGACATGGCC 3445
QY 235 -----AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal-----11 250
Db 3446 CTCAGAGAGATCCGGGAGCTGGAAATCTCAGATCTTGAACTCCAGGAGACCTGGAGTCT 3505
QY 250 eThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAl 270
Db 3506 GAGCGTCTTCCAGGAATAAAGCTGAGAGACAGAAACGGGACCTTGGGGAA-CAG----- 3559
QY 270 aPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSe 290
Db 3560 -----CTAGAGCGCGCTGAAAACAGAGATTGGAGGACACGCTGTGATTCACAGCTGCCCA 3612
```


Db 4748 AAGACGCGAGCTGGAAGAGCTGGAGGACGAGCTGCAGGCCA 4787

Search completed: July 12, 2005, 00:56:09
Job time : 1053.33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 23:52:39 ; Search time 14.6667 Seconds
(without alignments)
4073.905 Million cell updates/sec

Title: US-10-018-929C-3_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILGSSLNPSHDVKH.....SGNNKHFSGSSNISONAPDV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3207	99.7	2254	2 D86215	protein T6D22.14 [
2	163.5	5.1	1465	2 T23056	chromodomain helic
3	160	5.0	1738	2 C94507	hypothetical prote
4	152.5	4.7	1330	2 H89567	protein T08A9.1 [i
5	149.5	4.6	992	2 T46337	hypothetical prote
6	147	4.6	962	2 C71617	SERA antigen/papai
7	145	4.5	3187	2 JC5837	364K Golgi complex
8	145	4.5	3225	2 I52300	giantin - human
9	143	4.4	1130	2 T34081	hypothetical prote
10	143	4.4	1269	2 F84730	probable myosin he
11	142.5	4.4	1422	2 T18404	chromatin remodel
12	142.5	4.4	1961	1 A16231	myosin heavy chain
13	142	4.4	768	2 T02572	hypothetical prote
14	141.5	4.4	1392	2 A43336	microtubule-vesicl
15	140.5	4.4	1290	2 A55034	chromosomal protei
16	139.5	4.3	1429	2 S22695	retin - human
17	139.5	4.3	1829	2 T34239	hypothetical prote
18	139.5	4.3	1940	2 A29320	myosin heavy chain
19	139.5	4.3	3147	2 T18674	hypothetical prote
20	139	4.3	853	2 T51505	hypothetical prote
21	139	4.3	1388	2 S74245	serine/threonine-s
22	138.5	4.3	1927	2 A59236	embryonic muscle m
23	138.5	4.3	1939	2 T48175	myosin heavy chain
24	138.5	4.3	1972	1 A41604	myosin heavy chain
25	138.5	4.3	1999	1 S21801	myosin heavy chain
26	138	4.3	741	2 S39082	myosin heavy chain
27	138	4.3	3259	1 A56539	giantin - human
28	137.5	4.3	2116	2 A26655	myosin heavy chain
29	137	4.3	806	2 T08932	hypothetical prote

30	136.5	4.2	1662	1 H71402	probable kinesin -
31	135.5	4.2	784	2 D96564	myosin-like protei
32	135.5	4.2	1005	2 B87898	protein F10G8.8 [i
33	135.5	4.2	1041	2 T20730	hypothetical prote
34	135.5	4.2	1128	2 G86266	hypothetical prote
35	135.5	4.2	2168	2 T30171	ninein - mouse
36	135.5	4.2	3488	2 T34418	hypothetical prote
37	135	4.2	474	2 T50258	hypothetical coile
38	135	4.2	1186	2 C64588	cag pathogenicity
39	135	4.2	1295	2 T24587	hypothetical prote
40	135	4.2	1957	2 T38077	hypothetical coile
41	134	4.2	1558	2 B71603	RESA-H3 antigen pF
42	134	4.2	1927	2 G64585	cag pathogenicity
43	133.5	4.2	634	2 E86293	T24D18.1 protein -
44	133.5	4.2	750	2 T38435	coiled coil protei
45	133.5	4.2	824	2 T20477	hypothetical prote

ALIGNMENTS

RESULT 1

D86215
Protein T6D22.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D86215
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86215
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2254 <STO>
A:Cross-references: UNIPROT:Q9LN02; GB:A5005172; NID:98778840; PIDN:AAF79839.1; GSPDB:GN C:Genetics:
A:Gene: T6D22.14
A:Map position: 1

Query Match 99.7% ; Score 3207; DB 2; Length 2254;
Best Local Similarity 99.7% ; Pred. No. 3.3e-187;
Matches 619; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	RADAFILGSSLNPSHDVKHVEKIKIESCSERTKIFRLYSVCTVEERKALILARQNMQRK 60		
Db	883	RADAFILGSSLNPSHDVKHVEKIKIESCSERTKIFRLYSVCTVEERKALILARQNMQRK 942		
Qy	61	AVENLNRLTHALIMWCASYLFDKLDHFHSSETPDGSVFQSIQMDGVIHFFSILSKG 120		
Db	943	AVENLNRLTHALIMWCASYLFDKLDHFHSSETPDGSVFQSIQMDGVIHFFSILSKG 1002		
Qy	121	GEENEVKCLLLEAKHAQGYSSDSTLFGEDHIKLSDEESPNIWFKLLGKNPMWKYPS 180		
Db	1003	GEENEVKCLLLEAKHAQGYSSDSTLFGEDHIKLSDEESPNIWFKLLGKNPMWKYPS 1062		
Qy	181	DTPONRKRVOYFEGSEASPKTGDGNAKKRKKASDDVTPRVDDPPVDDDERKASGKH 240		
Db	1063	DTPONRKRVOYFEGSEASPKTGDGNAKKRKKASDDVTPRVDDPPVDDDERKASGKH 1122		
Qy	241	MGALESFKVITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 300		
Db	1123	MGALESFKVITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 1182		
Qy	301	QRRILHTVLPKPMKLCVLIHLSDACTSMVGNFLBYVTENHRIYEETPTTFOAFQIALSWI 360		

```
Db 1183 QRRLLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVIEHRIIYEPEPATTFOAQIALSWI 1242
Qy 361 AALLVVKQILSHKESLVRANSELAFKCSRVEVDYIYISILSCWKSLFLEHTQGLQDFCGTIN 420
Db 1243 AALLVVKQILSHKESLVRANSELAFKCSRVEVDYIYISILSCWKSLFLEHTQGLQDFCGTIN 1302
Qy 421 SKQSVSTKLAVNESLSCATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIIS 480
Db 1303 SKQSVSTKLAVNESLSCATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIIS 1362
Qy 481 GIKKKYKQVQKLQVQHEHEKMKELLNMYADKKQKLETSKVEAAVIRITCSRTSTQVGDIL 540
Db 1363 GIKKKYKQVQKLQVQHEHEKMKELLNMYADKKQKLETSKVEAAVIRITCSRTSTQVGDIL 1422
Qy 541 KLLDHNRYERFDEIKSEKNECLSKLEQMHVEVAKKLAEDACWINRIKSWAAKLKVCVPI 600
Db 1423 KLLDHNRYERFDEIKSEKNECLSKLEQMHVDVAKKLAEDACWINRIKSWAAKLKVCVPI 1482
Qy 601 QSGNNKHFGSGSNITSONAPDV 621
Db 1483 QSGNNKHFGSGSNITSONAPDV 1503

RESULT 2
T23056
Chromodomain helicase H06001.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T23056
R;Barlow, K.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19665
A;Accession: T23056
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1465 <WIL>
A;Cross-references: UNIPROT:O17909; EMBL:Z92970; PIDN:CAB07481.1; GSPDB:GNO0019; CESP:HQ
A;Experimental source: clone H06001
C;Genetics:
A;Gene: CESP:H06001.2
A;Map position: 1
A;Introns: 44/1; 91/3; 170/3; 377/3; 494/3; 1046/2; 1099/3; 1242/3; 1298/1
C;Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 5.1%; Score 163.5; DB 2; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.063;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

Qy 2 ADAPILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKA 61
Db 799 ADTVIIIFDSWNPNQDLQAMSAHRIQGTIVNIYRLVTGKSVEE-----EI 845
Qy 62 VENLRSLTHALLMWGASYLFDKLDH-----PHSSETPDGVSFQSGIMDVG 108
Db 846 VERAKRLV-----LDHLVIQRMDTTGKTVLSKNATASGVSVPFDQELSAI 891
Qy 109 IHFSSIT--LSSKGGENEV-----KCLLLEAKHAQGTSSDSTLPG--DHKLSDEES 160
Db 892 L-KFGAVELPKEREGEPEVDIDRLIMGAETREAEVEEYKGNELLSSPKYANFAIDEE 950
Qy 161 PNIF-----WSKLLGGKNPMKYPSPDTPQNRKRV-----Q 191
Db 951 KDIATAATDEAAII-----PEDRNRILEERMKELAEMLAPROKQIPQ 997
Qy 192 YFEGSEA-SPKTGGGNAKRRKKASDDVDTPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTCKKKKKAAGVNTTPEIKRFTKSPKFSMPLNRLLEEIAQDAELE 1057
Qy 235 ASGKDHWGALES PKVITLQSSCKSSGDTGLDNDAPGLYSMSGSHSIGIPEDMLASODWG 294
Db 1058 EHSTDEMKKL-----VESLSACKKADE--FDSNEKNG-----DAG 1092
Qy 295 KIPDESQ-----RRLLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYV 336
```

```
Db 1093 AAESKKQIERKPKFTCDVNLKQIERSHAELKP-----LHEILKSEETKTS----- 1139
Qy 337 IENHRIYEPATT-----FOAFQIAL-----SWIAALLVKQI--LSHKESL 375
Db 1140 -----FKPANAKLQKGDVDMRSRDPDSALLLGVWKYGYGSWEAKMDPTLGLADKIFI 1193
Qy 376 VRANSELAFKCSRVEVDYIYISILSCWKSLFLEHTQGLQDFCTGNSQSVSTKLAVNESL 435
Db 1194 KQTKPKQGGKNLQVRVDYLLKLM-----SKDKVKTTTE----- 1225
Qy 436 SGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIISGKKYKQVQKLVQ 495
Db 1226 -----KKERRKADDVPVGPPEKKKRHNNTVPQSEK-----KKKEK----- 1262
Qy 496 EHEKMKELLNMYADKKQKLETSKVEAAVIRITCSRTSTQVGDILKLLDHNRYERKFEIK 555
Db 1263 --EKNSSSLK---DQLALLSIDKSLYGGALEDSASAKPFLECVKLCMPVHKYMKLKEAQ 1317
Qy 556 SEKNECLKS-----LEQMHEVAKKLAEDACWINRIKSWAAKLKVCVCP 599
Db 1318 EAKNQADEAKYLRDGLDSFLENLETLIKKKPKTNIRKWNLYLWIFLCKFTLREP 1371

RESULT 3
C84507
Hypothetical protein At2g13370 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: C84507
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Neece, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84507
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1738 <STO>
A;Cross-references: UNIPROT:Q9SI41; GB:AE002093; NID:g4733988; PIDN:AAD28668.1; GSPDB:GN
C;Genetics:
A;Gene: At2g13370
A;Map position: 2
C;Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 5.0%; Score 160; DB 2; Length 1738;
Best Local Similarity 17.3%; Pred. No. 0.13;
Matches 132; Conservative 104; Mismatches 241; Indels 286; Gaps 28;

Qy 2 ADAPILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQN----- 56
Db 1041 ADTVIIFDSWNPNQDLQAMSAHRIQGTQVEVNIYRFTVTSKSVEEBILERAERKRVLDHL 1100
Qy 57 -----RONKAVERNLRSLTHALLMWGASYLFDKLDHFSSETPDGVSFQSG 103
Db 1101 VTQKLNAGLEKRETKGSGNFPDKNLSAILRFGABELFKE----- 1141
Qy 104 IMDGVIHFESSILSSKGGENEVKLCCL-----LEAKHAQGTSSDSTLGEDH 152
Db 1142 -----DKNDESKRLLSMDIDILERAQVQVEKH---TDETEHELLGAPK 1184
Qy 153 I-KLSDEESPNIFWSKLLGGKNPMKYPSPDTPQNRKRVQYFPGSEASPKTGGGNAKRR 211
Db 1185 VANFCAEDDGSFWSRWIKPDSVVTAEALAPAAANTKSYVDPSHPD-----RTSKRK 1238
Qy 212 KKASDDVDTPRVDDDERKASGKDHWGALES PKVITLQSSCKSSGDTGLDNDAP 271
Db 1239 KKGSE-----PPEHTERSQKREKTEYFVPSTPLLEGTSAQVRGWSYGNLPRKDAQ 1288
Qy 272 GLY-----SMGSH-----ISGIPEDMLASODWKIPDESQRRLHTVLKPKMAKLCQV 318
Db 1289 RFYRTVMKFGNHNQMACIAEVEGVVE-----AAPBEAQVELFDA----- 1328
```

```
Qy 319 LHLSDACTSMV--GNF-----LEYVNIHRIYEBPATTQ 351
Dy 1329 --LIDGCKESVETGNFEPKGPVLDFFGVPVKANELLKRVQGLQLLSKRSIRYNDFISQFR 1386
Qy 352 AFQIALGWIA-----ALLVKOILSH--KESLVRANGELAF--KCSRVEV 391
Dy 1387 -----VLSYLKPSNWSKGGWNQIDARLLGILYHGFNGWEKILRDSLSLGTKKIAPVEL 1442
Qy 392 DYIYSILCMKSLFLEHTQGLQPCF-----GTNSKQSVVSTKLNVNLSLG---ATVRDE 443
Dy 1443 QHHTFELPRAPNLKERATALLEMELAAAGKNTWAKASRKNSKVKVDNLINQFAPARDR 1502
Qy 444 K-----INTK-----SVNSSE-----DEE 458
Dy 1503 RGKSGPANVSLSTKDGPRKTQKAEPLVKEBEGMSDDGEVYEQFKEQKMWECEDVLADE 1562
Qy 459 CMTEKRCSHVSTATRD1--BKTISGIKKYK--KOVOKLVOEHEEKMLLNMYADKKQ 513
Dy 1563 IKTLGRLOQTTSADLPKKEVLPKIRRYLEILGRRIDAIVLEHEE-----DLYKDORM 1616
Qy 514 KLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFDTEIKSEKNECLSKLEQMHEVAK 573
Dy 1617 TMRL-----WNYVSTFNSLSDR-----LNQIYSLK 1643
Qy 574 KKLAEDEACWINRIKSWAAKLVKVCPIQSGNKNHFGSSNISQ 616
Dy 1644 QEKEEBEGV-----GPSHLNGSRNFOR 1665

RESULT 4
H89567
protein T08A9.1 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: H89567
R:anonymous, the C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: H89567
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1330 <STO>
A:Cross-references: UNIPROT:Q22342; GB:chr_X; PIDN:AAA81420.1; PID:g1065501; GSPDB:GN000
A:Gene: T08A9.1
A:Map position: X

Query Match 4.7%; Score 152.5; DB 2; Length 1330;
Best Local Similarity 19.0%; Pred. No. 0.26;
Matches 132; Conservative 112; Mismatches 225; Indels 227; Gaps 33;

Qy 58 QNKAIVENLRSL-----THALLMWGASYLFKDLHF-----HSSPTPSGVSFQSIM 105
Dy 446 EERSREQLGRKLKXHLHALF-----PGLFDNLPEFYVKAPLEKYDTDLPLNAKEYIKDLR 501
Qy 106 DGV-----IHFESSILSSKGGENEVKLLLEAKHAQGYTSSDSTLFGEDHIK- 154
Dy 502 DALPELEFFLKVLPNVAASKLTKGSKQH-----PSFGQTRVES 540
Qy 155 -LSDEE-----SPNIFWSKLGGKNP-----MWKVPSPDFPQNRKR-----VQ 191
Dy 541 FLTDEPMRMARSHNYNPAWLSEDDGSSFPVPOPLMCRSPENALRESQCRSIPYVPSLQ 600
Qy 192 YFEGSEASPTGDCGNNAKRRKASDDVTDPRTVPPVD--DDERKASGKD--HMGAL--- 244
Dy 601 QLEGLD-GRAPG-----SSAPISIPNSTSSRINFKSGRQSSQDLHVGSAVSS 649
Qy 245 -----ESP-----KVITLQSSCKSGTGDTLGDNDAP-GLYSMGSHSIGIPED-MLAS 290
Dy 191 -----KVITLQSSCKSGTGDTLGDNDAP-GLYSMGSHSIGIPED-MLAS 290
```

```
Db 650 DTSLLGHETPVKMEDTVLMTENKSELNLSNMNDSIDSLLDVYEHIDYSAEDILLGL 709
Qy 291 QDMGKIPDESQRRLHTV-----LKPKVAKLCQVLHLS-----DACTSMVGNF 332
Dy 710 SDHTVQLEDAVKRLSKMQNIVANMTPRNEKLFKYLHQTPAPRDFEIIIHGDKLKNRVKEL 769
Qy 333 -----LEYVNIHRIYEBPATTQAFQIALSWIAALLVKOILSHKESLVRANSELAFKC 386
Dy 770 GESVESLELENKCKSTE-----SNLEAKLAEMTVSHHKEIENTOAE----- 811
Qy 387 SRVEVDYIYSILCMKSLFLEHTQGLQPCF-----GTNSKQSVVSTKLNVNLSLGATVRDEKIN 446
Dy 812 -----CIKMSVEFE--LMTDSMSRQSKQIESKDEIRELKALEKQNTIAH 856
Qy 447 TKNMRNSEDDECMTEKRCSHYSTATRDIEKTIISGIKKYKQVOKLVQ----- 495
Dy 857 EKALRNDPYSEE-----YKRTLTAIR---AELEKEFKQRIEVTITKAVECKDEAFA 905
Qy 496 -----EHEEKQMLLNMYADKKQKLET---SKSVEAAVI-----RITCS 531
Dy 906 ROEKTLEIENRVLSSENESSKKEANMR---EKEQLESIRIQMPEGEVILDEFNALKES 962
Qy 532 RTSTQVGDLL-----KLDHNYERKFDTEI-KSEK-NECLSKLEQMHEVAKKK 575
Dy 963 QPSAQIKDRFAAIRSRMERSTDGKCLDTNKLVDNLKLVKHLQQLVNEREQYRDVIKQ 1022
Qy 576 LAEDEACWINRIKSWAAKLVKVCPIQSGNKNHFGSGS 611
Dy 1023 PGGDQV--LQFLENSTPTVPVDM EAVFELTNFSGS 1056

RESULT 5
T46337
hypothetical protein DKFZp43402413.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46337
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223037
A:Accession: T46337
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-992 <AAA>
A:Cross-references: UNIPROT:Q9NTH6; EMBL:AL137265
A:Experimental source: adult testis; clone DKFZp43402413
C:Genetics:
A:Note: DKFZp43402413.1

Query Match 4.6%; Score 149.5; DB 2; Length 992;
Best Local Similarity 19.0%; Pred. No. 0.27;
Matches 125; Conservative 98; Mismatches 219; Indels 215; Gaps 26;

Qy 122 ENEVKLLLEAKHAQGYTSSDSTLFGEDHIKLSDEESNIPWSKLGG----- 171
Dy 91 EKSEPKICRNLVTPKADPT-----GSEPAKASEKAPE---DTVDAGEGSRREAA 139
Qy 172 KNPWMKY-----PSDTPQ-----RNRKRVQYFEGSEASPKTGDGNAKRRKKASDDVD 220
Dy 140 KEPAKASALEEGSSDASQBLEISEHMKFQLSDSTASDPKSPHGLDFGFRSRISEHLLD 199
Qy 221 PRVTDP-----PVDDDERKASGKHMGALES-----PKVITLQSSC 256
Dy 200 VDVLSPLVGGACRAQAPGLIGIEDKDSQSSQDELQSKQSGLEERLSPPLPHEERAQSP 259
Qy 257 KS-----SGTDGTLN-----GNDAFGLYSM----- 276
Dy 260 RSLATEEPPOGPEGOPFWEKABELGDSAAISLSQLSLQREQAPSPPAACEKGEHQSQ 319
Qy 277 -----GSHISGIPEDMLA-----SODWGKIPDE 299
Dy 320 ABELFGQEBEADPEEKVAVSPPTPVSPVSTPEVPAPPEQLSEALKAMEEVAQVLEQ 379
```

QY 300 SORRLHTVLKPKM-----AKLCQ-----VLHL-----SDACTSMVGNFLEVYIENHRIYBEPA 347
Db 380 DQHHLESKEKQWOLREKLCQEBEEIILRHQKEQSLSLRERLQKAIE-----EEEA 434
QY 348 TTFQAFQIALSWIAALIVKQILSHKESLVANSFLAPKCSRVEVDYIYISLSCMK-----S 403
Db 435 RMREESQRLSWLRA-QVQSSTQADEQIRAEQASQLKLEELE-----SQQKAERAS 487
QY 404 LPLEHTQGL-----QPCFGTNSKQSVST-----KLVNESLSGATVRDEKINTKSMRN 452
Db 488 LEQKNQMLBQLKEIEA-SEKSSQALNAKAELAQQLREQLGE--RKEAVATLEKH 544
QY 453 SSBDEECWTEKRCSHYSTARDI-----EKTISGTKKKYKQVQKVLQVEHEEK----- 500
Db 545 SAE-----LERLCSLEAKIREVVSSLOKKIQEAAQQAQKLCQGVHRVHVKSYHV 599
QY 501 ---KMELLNMYADKKQKLE-----TSKSVEAAVIRITCSRTSTQV 537
Db 600 AGVEHELSSLREKQVEGEHERRLDKMKEEHQOVMAKAREQVEAEERKQRAELLGHLT 659
QY 538 GDLKLLDHNRYERKDEIKSEKNECKLSLEQMHVEVAKKLADEACWINRIKSWAAKL 594
Db 660 GELERLQRAHERLETVRQGHKLEDLRRHREQERKQLQDLLELSTRAKDVKARL 716
RESULT 6
C71617
SERA antigen/papain-like proteinase with active Ser PFB0345c - malaria parasite (Plasmid
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: C71617
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perceca, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: C71617
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-962 <GAR>
A;Cross-references: UNIPROT:O96164; GB:AE001388; GB:AE001362; NID:G3845157; PIDN:AACT185
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0345C
C;Superfamily: Plasmodium vivax serine-repeat antigen
Query Match 4.6%; Score 147; DB 2; Length 962;
Best Local Similarity 20.3%; Pred. No. 0.36;
Matches 101; Conservative 75; Mismatches 172; Indels 150; Gaps 28;
QY 194 EGSEASPKTGGGNVAKRKKASDVTPRVTDPPVDD-----DERKASGRDHMGALE 245
Db 43 EGSDEVSGAGQDASQAGQDANPDPTLPKPPSPADPTKDTGSGQDADSSSSKTEIPLV 102
QY 246 SPK-----VITLQSSCKSSGDTGLDGNDAFLYSMGs----HSGIPEDMLASODWGKIP 297
Db 103 KPNHKTIVAMLNKYNKGVKVTGTC-GAD-FGLFLPHIYVHVKSEDTIELSELAA--P 158
QY 298 DESQRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYIEN-----HRIYEE----- 345
Db 159 PEMQTKFD---KTQLKFCF---VKDDTKKF--DFIAYIVKDIILVFKWKVYEGLSKEQD 209
QY 346 -----PATTFOAFQIALSWIAALLVQKILSHKESLVANSFLAP-----KC 386
Db 210 VDEMKYLLPNLKQPIITSIQHV-----SWTGT-----KESYTLSESKDYVLGEGMPEKC 256
QY 387 SRVEVDYIYS-----ILSCMKSLFLEHTQGLQFQDCF---GTNSKQSVSVSTKLVNESLSGAT 439
Db 257 DATATDCFSGFTDYGKFCQCKLLMQEKNINDSCFKYVSSNQKELI----- 302
QY 440 VRDEKINTKMRNSESDEECWTEKRCSHYSTATRDIEKTIISGIKKYKQVQKVLQVBEHE 499
Db 303 -----KKQLKITAQDDSESSE---YHLSSEIKNLLKNI-----YKK-----NNDDN 340

QY 500 KMELL---NMVADKKQKL-----ETSKSVEA-AVIRITCSRTSTQVGDLLKLLDHNRYERKF 551
Db 341 KKELHLHFENVSALSKELLNLYCNLLKEVMNGVLK-----DHQLVQDVFNNTLKL 394
QY 552 DEIKSEKNECKLSLEQMHVEVAKKLADEACWINRIKSW-----AAKLKVCVPIQSG 603
Db 395 EEHKEENDNVL-----YHKM-----KNEALCLKVNDWMKNKTGLLLPQLSYDLTYKNN 443
QY 604 NNKHFS-GSSNISONAPD 620
Db 444 NFTEFTQNKSYTSQNIVD 461
RESULT 7
JC5837
364K Golgi complex-associated protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C;Accession: JC5837
R;Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A;Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
P;45-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predicted
F;3165-3187/Domain: membrane anchor #status predicted <MAD>
A;Reference number: JC5837; MUID:98093490; PMID:9431462
A;Accession: JC5837
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3187 <TOK>
A;Cross-references: UNIPROT:Q63714; DBJ:D95543; NID:9516825; PIDN:BA05026.1; PID:95168;
C;Comment: This protein plays a role in the formation and maintenance of the characteris
C;Superfamily: Glantin
Query Match 4.5%; Score 145; DB 2; Length 3187;
Best Local Similarity 20.7%; Pred. No. 2.4;
Matches 142; Conservative 97; Mismatches 244; Indels 202; Gaps 35;
QY 1 RADAFILFGSSLNPSHDVHKVHKIKIESCSERTKIFRLYSVCTV--EEKALIILARQNM 58
Db 314 QAELEWQYQ-TLQORHETEMEEKTACISLLQKNE-QELQSACDALKEENSKLLQEQQEA 371
QY 59 NKAIVENLRSLTHALLMWCASYLFDKLD-----LEDELQKSKSISQFVNKPNLEKHETSSQTSLSLPVYNE 95
Db 372 AKSAQALQ-----LEDELQKSKSISQFVNKPNLEKHETSSQTSLSLPVYNE 418
QY 96 -----SGVSPESQIMDGVIIHFPSSILSSKGG-----ENVKLC---LLLAKAKAQGT 140
Db 419 GVQAVMEESVASIQKRVLENEKGAALLSSLEELAEAE-KLCSRITLLEAQNRAE 477
QY 141 YS-----SDST-----LFGEDH-----IKLSDEESPNI FWS 166
Db 478 ADGMVCEVSTAGTALINRSDSSTESGQDVLENTFSQKHKELSVLLVEMKEAQEEIAFLK 537
QY 167 KLGGGNPKMVKYSDPTFQRNKRQVYFEGSEASPFTGDGNGAKRKKASDVDPDPRVTD- 225
Db 538 SQLGKRP-----EGDYEVLDRKEVQWME-SEG LPSV-----TARDVLCAPRDKNS 582
QY 226 -PPVDDDERKASGDHMGALSPKVITLOSSCKSSGDTGLDGNDAFLYSMGSHISGIP 284
Db 583 VPAVEGEQ--AGMRDQHTLEAGPLNDTGMELNSPQPDGVDKLSA-----P 627
QY 285 EDMLASQDWGKIPD-----ESQRRRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVIE 338
Db 628 HVCLCHQ--GELERLKTQVLELETSLHTAKETHEKNLSEKAKKEISSITQTKEFKESA-- 683
QY 339 NHRIVSEPARTFOA-----FOIALSWIAALLVQKILSHKESLVANSFLAPKCS 387
Db 684 -----BEARSTLTAVERDQLLYRVKELDVGLRAQVRELETSLAEAKQKRGLDVYESQ 738
QY 388 RVEVDYIYSILSCMKSLFLEHTQGLQFDCFGTNSKQSVSVSTKLVNESLSGATVR--DEKI 445
Db 739 RAQ-----HNLITQIHSLSIE-----AKSKDVKIETLQRELDGVQLQFSQGT 782

Db 854 DDWTOQLNEBTTVVLPDINSIQEKIDEKEATINEMNERLKSGRE--NEIAK 900

RESULT 10

F84730

probable myosin heavy chain [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: F84730

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: F84730

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1269 <STO>

A;Cross-references: GB:AE002093; NID:g6598483; PIDN:AAC69932.2; GSPDB:GN00139

C;Genetics:

A;Gene: At2g32240

A;Map position: 2

Query Match 4.4%; Score 143; DB 2; Length 1269;

Best Local Similarity 21.9%; Pred. No. 0.92; Indels 120; Gaps 20;

Matches 110; Conservative 81; Mismatches 192;

Qy 150 EDHKLKLSDESPNIFWSKLGGKNNPKYPSDTPQNRKRQVQPEGSEASPKTGDGNAK 209

Db 2 EEAQTTSSEVP-----VVKGRQRCYCLDLLQAVNVEPKKEE-----DGFIK 50

Qy 210 KRKASDDVTPRVTDPVDDDERKASGDHMGALSPKVIITLQSSCKSGTDTLGDND 269

Db 51 VEKEAFD-----AKDDAEKA--DHVPVEQKEVI-----ERSSSGSQRELHESQ 92

Qy 270 -----AFGLYSNGSHISGIPEDMLASQDNGKIPDESQRRLHTVLPKMAKLC 316

Db 93 EKAKELELERVAGELKRYESENTHLKDELLSAKEKLETEKHGGLVQVKKQEKIV 152

Qy 317 QVLHLSDACTSMVGNFLEYVIEHRIYEEPAT-----TFQAFQIALSWAALLV----- 365

Db 153 E----GEERHSSQLSLEDALQSHDAKDKEITVEKFAFDALGIELESSRKLIIELEBGLK 208

Qy 366 -----KQILSHKESLVRANSELAFKCSRVEVDYIYISLCKMSKL--FLEHTQGLQFDC 416

Db 209 RSABEAQKPBELHKQSAHADSE-----SOKALEFSELLKSTKESAKEMEERKASLQOEI 263

Qy 417 FGTNSKQS--VYSTKLAVNESLSGATVRDEKINTKSMRNSSE-----D 456

Db 264 KELLNEMSENEKVEAALKSSAGELAAVQBELALSLSRLLETEQKVSSTEALIDELTQELE 323

Qy 457 EECWTEKRCSHYSTATRDIEKTISGIKKYKQV---QKLQVHEEKKMELLNNYADKKQ 513

Db 324 QKKASESRFEELSVLDLQAOQTGLQAKLSEQGGINSKLAELKEL--LESLSKDQEE 382

Qy 514 KLSTSS-----KSVBAVIRITCS--RTSTQVGD-----LKLDDHYERKFDI-- 554

Db 383 KLRTANEKLAELVKEALEANVAEVTSNVATVTEVCNLEELKLTSDENFS--KTDALLS 441

Qy 555 -----KSEKNECKLSLEOMHEVA 572

Db 442 QALSNNSELQKLSLSELHSEA 464

RESULT 11

T18404

chromatin remodelling complex protein SNF2L - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum

C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004

C;Accession: T18404

R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, R.; Ji, D.D.; ARNOT, D.E.

A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different

Mol. Biochem. Parasitol. 88, 151-162, 1997

A;Title: A Plasmodium-falciparum homolog of the ATPase subunit of a multi-protein complex

A;Reference number: Z09426

A;Accession: T18404

A;Status: preliminary; translated from GB/EMBL/DBSJ

A;Molecule type: mRNA

A;Residues: 1-1422 <JID>

A;Cross-references: UNIPROT:O00914; EMBL:AF003086; NID:g2213861; PID:g2213862; PIDN:AAC47

Query Match 4.4%; Score 142.5; DB 2; Length 1422;

Best Local Similarity 18.8%; Pred. No. 1.1;

Matches 123; Conservative 111; Mismatches 241; Indels 187; Gaps 29;

Qy 2 ADAFILGSLNPSHDVQKHVEKIKIESCBERTKIFRLYSVCTVEEK-----AL 49

Db 711 ADIVILFSDYNDPQMDIQMDRAHRIGQKKRVIVYFVTQNSVEEKIVERAAKKLKDLSL 770

Qy 50 ILARQNRQKAVENRSLTHALLMWGASYLPDKLDHFHSS-----ETPDSGVSP 100

Db 771 IQQKGLNLNSAKEN-NKQELHDILNFGAPEVYKTDQISSISDESDIILADAERKTIEI 829

Qy 101 EQSI-----MDGVITHEPSSIL-----SSKGEENEVKLCLLEAK 135

Db 830 EKKLKNLENIFDLTNTISLQGLNMYNDELKEASEESTDEEDSSGSEET-----ILEG- 882

Qy 136 HAQGTYSSTLTFGEDHIKLSDESPNIFWSKL-----LGGKNPMW 176

Db 883 -----SGNDIVEEVNGVKKKKKKNI--NKIRRIKFLKNNKNNKMTFLDGERKSKW 934

Qy 177 KYPSTDPQRNRKRQVQPEGSEASPKTGDG---GNAKK---RKASDDVTPRVTDPVDD 230

Db 935 KVNTACGRNTKKKVVLLHGWRAEARGHDFQFFNVEKLDLEKIEDKWNMYMINQEKINV 994

Qy 231 -DERKASGDHMGALSPKVIIT-----LQSSCKSGTDTLGDNDAGF-----LYSM 276

Db 995 LIEAQNEEKDY-----EKIVTFNEFLDKHAKNIYQIITLINPSLDFEHKDTKEELLKM 1048

Qy 277 GSH--ISGIPED---MLASQDW-----GKIPDESQRRLHTV-----L 308

Db 1049 GTHSDVNSNDKDYKRSFSFSTDYNNKTYLNNNSNNNNNNNNNDNVGTSINTIDQQYF 1108

Qy 309 KPMK-----AKLCQVLHLSDA---CTSMVGNFLE-----YVIENHRIYEEPATTFQAFQ 354

Db 1109 KKKILSPFESGKKLQMIKPKDRNIKCYQTQVANFIKKNFPEKINNNDNNKSTTLKNKK 1168

Qy 355 IALSW-----TAALLVKQILSHKESLVRANSELAFKCSRVEVDYIYISLCKMSKL 404

Db 1169 LKKGKQTKNNVNEKDIASVDIEKIKLQKQELMKQGFA---KWNKAEFNKL-----MSG 1219

Qy 405 FLEHTQGLQ--FDCFGTNSKQSVVSTKLAV-----NESLSGATVRDEKINTKSMRNSSE 455

Db 1220 IYGTNEVEYIYEKYSNKSMSMEDIKAYLTVFPRKYDQIKGGVRLFDKIKRSDLOKKII 1279

Qy 456 DEECWTEKRCSHYSTATRDIEKTIS-GIKKKYKQVQKLQVHEEKKMELLNNYADKKQ 514

Db 1280 EEE-----NDMITEFVEKQLSEGVDSIEKQLPSNRLRYDFMEKREVLNITEEVKTE 1330

Qy 515 LE 516

Db 1331 NE 1332

RESULT 12

A61231

myosin heavy chain nonmuscle form A - human

N;Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A

N;Contains: myosin ATPase (EC 3.6.4.1)

C;Species: Homo sapiens (man)

C;Date: 12-May-1994 #sequence_revision 14-Jul-1994 #text_change 09-Jul-2004

C;Accession: A61231; A34876; I52562; I61692

R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, R.; Ji, D.D.; ARNOT, D.E.

A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different

A;Reference number: A61231; MUID:91316803; PMID:1860190

A;Accession: A61231
A;Molecule type: mRNA
A;Residues: 1-715 <SM>
R;Saeg, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A;Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through alternative splicing
A;Reference number: A34876; MUID: 90138958; PMID: 1967836
A;Accession: A34876
A;Molecule type: mRNA
A;Residues: 715-1961 <SAE>
A;Cross-references: GB:M31013; NID: g189035; PIDN: AAA36349.1; PID: g189036
R;Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.; Blood 78, 1826-1833, 1991
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, cDNA sequencing, and expression analysis
A;Reference number: 152562; MUID: 92003925; PMID: 1912569
A;Accession: 152562
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-52, 'EAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>
A;Cross-references: GB:M81105; NID: g188988; PIDN: AAA59888.1; PID: g553596
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Moosker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A;Title: Identification and overlapping expression of multiple unconventional myosin genes
A;Reference number: A55758; MUID: 94294418; PMID: 8022818
A;Accession: 161692
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 182-218 <BEM>
A;Cross-references: GB:L29141; NID: g457249; PIDN: AAA20904.1; PID: g531134
C;Genetics:
A;Cross-references: GDB:120216; OMIM: 160775
A;Map position: 22q12.3-22q13.1
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide binding site; myosin motor domain homology <MOTOR>
F;84-764/Domain: myosin motor domain homology <MOTOR>
F;174-181/Region: nucleotide-binding motif A (P-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1938/Domain: coiled coil #status predicted
F;837-1277/Domain: S2 #status predicted <DS2>
F;1278-1961/Domain: light meromyosin #status predicted <LM>
F;1939-1961/Domain: carboxyl-terminal #status predicted
F;125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694, 704/Active site: Cys #status predicted

Query Match 4.4%; Score 142.5; DB 1; Length 1961;
Best Local Similarity 18.8%; Pred. No. 1.8;
Matches 137; Conservative 104; Mismatches 216; Indels 273; Gaps 27;

Qy 22 EKIKSCSERTKIFRLYSVCTVEEKLILARQNMQRKAVENLRSLTHALLMWSYL 81
Db 960 QKQLEKVTTEAKUKL-----EEQIILEDQNCKLAK----- 992

Qy 82 FDKLDHFSSETPDGVSFQSIIMDGVIFHSSILSKSGGEENEVKLCLEAKHAQGT 141
Db 993 -----EKKLEDRIAEFTNLT-----EEKSKSLAKLNKH----- 1025

Qy 142 SSDSTLFGEDIHKLSDSESNIFWSKLLGGKPNWKPSTPQNRKRQVFGESASP 201
Db 1026 -----EAMITDLE-----ERLRREKQRELE----- 1047

Qy 202 TGDGNAKPKKASDDVTPVTPDDDERKASGDHMGALSPKVTITLQSSKSGT 261
Db 1048 -----KTRRLKLGSDTD-----LSQIAELQAIKLNKLAKEE-----ELQAAALARVEE 1094

Qy 262 DGLDGDNAFGLYSGWGHISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMKLCQV 321
Db 1095 EAAQKNWALKKIRLESQISELQEDLESERASRNKAQKQKEDGEEELAKTELED 1154

Qy 322 SDACTSMVGNFLEVYENHRIYEPATTFQAFQI-----ALSWAALL----- 364

Db 1155 TAAQQLRSKREQEVNLTNKKTLBEEAKTHEA-QIQEMRQKHSQAVEELAEQLEOTKRVKA 1213
Qy 365 -----VKQIL-----SHKESLV-----RANSELAF 384
Db 1214 NLEKAKOTLENERGELANEVKVLQGGFDSHKKKVEAQLOELQVKNFNEGERVTELAD 1273
Qy 385 KCSR--VEVDYIYSILSCMSLP-----LEHTQGLQFDCFGTNSKQSVVSTK 429
Db 1274 KVTKLQVELDNVTGLLSQSDSKSKSLTKDFSALESQLODTQELIQE---ENRQKLSLSTK 1330
Qy 430 LVNESLGSATVREKINTKMRNNSSEDE-----CMTEKRCSHYSTATRTIEK----- 477
Db 1331 L-----KQVEDEK-----NSFRQLEBEEBAKHNLKQIATLHAQVADMKKKWEDSVGC 1380
Qy 478 --TISGIIKKYKQVQKLVQEHBEK-----KME-----LNNM----- 507
Db 1381 LETAEVKKLQKDLGLSQRHEEKVAAYDKLEKTKRLQOELDLLVDLDHQRQACNL 1440
Qy 508 -----YADKKQKLET-----SKSVEAAVIRITCSRTSTQVGDLLKLD 544
Db 1441 EKQKQPDQLLBEKTTISAKYAEERDRAEAAREKETKALSARALEEAMQKAELERLN 1500
Qy 545 HNYERKDEIKSEKNECKLSLEQWHEVAKKLADEACWINRIKSWA-----A 592
Db 1501 KQFRTEMDLMSKDDVGKS---VHELEKSKRALEQQ---VEEMKTQLEEEDELEQATEDA 1555
Qy 593 KLKVCVPIQS 602
Db 1556 KLRLEVNLA 1565

RESULT 13
T02572
hypothetical protein At2g39300 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein T16B24.6
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C;Accession: T02572; F84815
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, D.; Nietman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID: 20083487; PMID: 10617197
A;Accession: T02572
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-768 <ROI>
A;Cross-references: UNIPROT:O80951; EMBL:AC004697; NID: g3402671; PID: g3402677
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nietman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID: 20083487; PMID: 10617197
A;Accession: F84815
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-768 <STO>
A;Cross-references: GB:AE002093; NID: g3402677; PIDN: AAC28980.1; GSPDB: GN00139
C;Genetics:
A;Gene: T16B24.6; At2g39300
A;Map position: 2
A;Introns: 80/2; 665/3

Query Match 4.4%; Score 142; DB 2; Length 768;
Best Local Similarity 19.2%; Pred. No. 0.55;
Matches 107; Conservative 94; Mismatches 201; Indels 154; Gaps 25;

Qy 84 KLDHFSSETPDGVSFQSIIMDGVIFHSSILSS-----KGGNEEVKL-----CLLL 132
Db 25 KDDHFQRYLNSPKGLNKSQSEVGAAUARRSRSSAAAFVDTGSSNOHRLNHSRCLTP 84

Db 699 KPAF-----YFALRD---TIVANNLD-----QATRVAFQDKRWVVTLQOI 738
Qy 369 LSHK-----ESLVRANSELAPKCSRV-----EV 391
Db 739 IEQSGTMTGGGKVMKGRMGSSVMVEISDDQLQKMENKLTDTTTRATEIQDKAHLEEV 798
Qy 392 DYIYSILSCMKSLFLEHTQGLQDFCGTNSKQSVVSTKLIVNESLSGATVRDEKINTKSMR 451
Db 799 AKLROATREMKNTPFKYTASLQ-----SLSEQEVHLK-----AQVKELEVNAAAA 844
Qy 452 NSSDEECWTEKRCRSHYSTATRDIEKTIISGIGIKYKQOVQLVQ-EHEEKKMELL---N 506
Db 845 PDKNOQ-----KQMEKNLETLLKKEKVAEKAGKVEAEVKRLHKLIVDIN 889
Qy 507 MYADKKQ--KLE-TSKSVEAAVIRITCSRTSQVGDGLKLLDHNVERKPFDEIKSEKNECL- 562
Db 890 NHLKAAQQDKLDKVTKEIDECASAITKAQVS-----IKTADRNLLKSEEAVARTEKEIVA 944
Qy 563 --KSLEQMHVAKKKLADEACWINRIK 588
Db 945 NDKSIEELTE-DLKKLEEKATTVMNECK 971

Search completed: July 12, 2005, 01:03:30
Job time : 20.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 10:23:56 ; Search time 5377.33 Seconds

(without alignments)
4395.843 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILFGLSLPNSHDVKH.....SGNNKHFGSSNISONAPDV 621

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Zgapop 6.0 , Zgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-O=/cgm2_1/USPTO.spool/ROBINSON929/runat_11072005_102152_29186/app_query.fasta_1.2325

-DB=EST -QPMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0

-UNITS=dib -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCFMGT=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=ROBINSON929 @CGN 1.1 8678 @runat_11072005_102152_29186 -NCFU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hcc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gsl1.*

9: gb_gsl2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	413	12.8	654	8	BH989471
2	327.5	10.2	775	9	CU818501
3	317.5	9.9	764	9	CU795819
4	299.5	9.3	369	8	BH419713
5	297.5	9.3	798	9	CG167203
6	227	7.1	525	4	BG833433
7	224.5	7.0	569	2	BZ203168
8	220	6.8	482	4	BG833434
9	214	6.7	668	5	BP133890

10	213.5	6.6	246	9	CC794080
11	211.5	6.6	516	1	AU288479
12	207.5	6.5	486	1	AI947519
13	200	6.2	799	8	B21272
14	198	6.2	560	8	BH788688
15	192	6.0	669	6	CA215647
16	183	5.7	942	8	B08967
17	181	5.6	529	5	BQ873535
18	179	5.6	969	9	CG396389
19	178.5	5.6	519	1	AU302228
20	168	5.2	690	8	BH978547
21	143.5	4.5	928	8	B11993
22	143.5	4.5	2265	3	AK014630
23	142.5	4.4	2079	3	AK017709
24	140.5	4.4	341	6	CB084203
25	140.5	4.4	3927	9	CL971593
26	140	4.4	2176	3	AK016726
27	139.5	4.3	1338	7	CO635655
28	138	4.3	2019	9	CL958706
29	137.5	4.3	3738	9	CL972961
30	136	4.2	3887	3	AK081567
31	135.5	4.2	3180	9	CL961241
32	135	4.2	4137	9	AY420057
33	134.5	4.2	848	8	BZ998864
34	134	4.2	1512	9	CL959694
35	133.5	4.2	3807	3	AK029994
36	133	4.1	2867	3	AK028516
37	132.5	4.1	1105	7	CR754231
38	132	4.1	668	6	CB526089
39	132	4.1	778	6	CA315071
40	132	4.1	2615	3	AK044987
41	132	4.1	5042	3	HS0805132
42	132	4.1	7050	9	AY420513
43	131.5	4.1	3344	3	BC028681
44	131	4.1	2133	3	AK014977
45	131	4.1	3336	3	AK029534

ALIGNMENTS

RESULT 1	BH989471	oem85a08.b1	B.oleracea002	Brassica oleracea	genomic, genomic survey
LOCUS	BH989471	654 bp	DNA	linear	GSS 07-OCT-2002
DEFINITION	BH989471	sequence.			
ACCESSION	BH989471				
VERSION	BH989471.1	GI:23526627			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
REFERENCE	Delehaency,K.,	Fewell,G.,	Fulton,L.,	McCombie,W.R.,	Miner,T.,
AUTHORS	Nash,W.,	Rabinowicz,P.D.	and Wilson,R.K.		
TITLE	Whole genome shotgun reads from Brassica oleracea				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Richard K. Wilson				
	Genome Sequencing Center				
	Washington University School of Medicine				
	Email: submissions@watson.wustl.edu				
	Plate: oem85 row: a column: 08				
	Seq primer: -21UPpOT forward				
	Class: shotgun				
	High quality sequence start: 20				
	High quality sequence stop: 551.				
	Location/Qualifiers				
	1. .654				
	/organism="Brassica oleracea"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:3712"				
	/clone_lib="B.oleracea002"				

FEATURES

source	
1. .654	
/organism="Brassica oleracea"	
/mol_type="genomic DNA"	
/db_xref="taxon:3712"	
/clone_lib="B.oleracea002"	

/notes="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassaica oleracea T01000DH3 buds, provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Alignment Scores: 1.27e-34 Length: 654
 Pred. No.: 413.00 Matches: 90
 Score: 413.00
 Percent Similarity: 67.95% Conservative: 16
 Best Local Similarity: 57.69% Mismatches: 28
 Query Match: 12.84% Indels: 22
 DB: 8 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x BH989471 (1-654)

Qy 358 SerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArg 377
 Db 253 AGTTGGATTGACGCTCGCTGTAAGCAACGTTAAACCGCGAAGAAATCTTTGGTCCGC 312
 Qy 378 AlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIle 397
 Db 313 GCAAAATCGGAATTAGTTTCAGTTGCTCTGAGAGAGGTAGATTATATATCTCTT 372
 Qy 398 LeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPhe 417
 Db 373 TTGTGTCGATGAAGAGTATATCTCTGGGGCGCACACGAAGTTTGCAGGTGGATTACTTG 432
 Qy 418 GlyThrAsnSerLysGlnSerValValSerThrLysLeuValLysGlnSerGly 437
 Db 433 GGTAATAATTCTAAGCAAGCA-----AAGGTGAATGAAGTCTCTCAGGG 477
 Qy 438 AlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGlu 457
 Db 478 GATACAGTGGCTCAGGAACAGAGTAAACAAATCGATCGT----- 519
 Qy 458 GluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLys 477
 Db 520 -----AGCTCAGTAACAAAGATGTGCAAAAG 546
 Qy 478 ThrIleSerGlyIleLysLysTyrLysGlnValGlnLysLeuValGlnGluHis 497
 Db 547 ACTATTAGCGACATCAAAAATAAATGCAAGTAAAGAACTGCAGGAGCTTGTAACAAATCCAC 606
 Qy 498 GluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGln 513
 Db 607 GAGGAAAAAAGGTGGATCTGTGATGATAGGAGTCTGACAAAGACAG 654

RESULT 2

CL818501 775 bp DNA linear GSS 09-AUG-2004
 LOCUS OR_CBa0036H03.r OR_CBa Oryza rufipogon genomic clone OR_CBa0036H03
 DEFINITION 3," genomic survey" sequence.

ACCESSION CL818501

VERSION CL818501.1 GI:51064111

KEYWORDS GSS.

SOURCE Oryza rufipogon

ORGANISM Oryza rufipogon

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 775)

AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.

ONAP project

Unpublished (2004)

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: http://genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0036 row: H column: 03

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

Location/Qualifiers

source 1..775

/organism="Oryza rufipogon"

/mol_type="genomic DNA"

/db_xref="taxon:4529"

/clone="OR_CBa0036H03"

/tissue_type="young leaves"

/dev_stage="2 week old seedlings"

/lab_host="DH10B T1 phage resistant"

/clone_lib="OR_CBa"

/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;

drx treated 36 hrs before harvest"

ORIGIN

Alignment Scores: 4.78e-25 Length: 775
 Pred. No.: 327.50 Matches: 85
 Score: 327.50
 Percent Similarity: 51.19% Conservative: 44
 Best Local Similarity: 33.73% Mismatches: 104
 Query Match: 10.18% Indels: 19
 DB: 9 Gaps: 2

US-10-018-929C-3_COPY_650_1270 (1-621) x CL818501 (1-775)

Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22

Db 6 GATGCCATCATCATATATGTTCTGCACACACCCCACTAACCGATCGATGAGGCTCTTCAA 65

Qy 23 LysIleIstIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42

Db 66 AAGATCAAGATAGAGTCACAAATTTGACGGCTGTGAGCAATTTTCGCTGTGACACTCTCTTT 125

Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62

Db 126 ACAGTTGAGGAGAGAGAGTCTTGTCTTCGAGGAGGAGGATTTATTGATTAACAACATA 185

Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82

Db 186 CAAGATCTAAGGACTAGTTTAAACATTCCTTACTCAGATGGGTGCTGCAATTTCTCTTC 245

Qy 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102

Db 246 AGTAGACTTGATGAGTCCAGCAAGATGACCATGCAAGTAAAGTTCTGAAATGGAGAGA 305

Qy 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlu 122

Db 306 CACTTATTGATGAAGTAACTCGTTGAGTTCTTAAACAAGTTATCCACAACCTGTTGAAGAC 365

Qy 123 GluAsnGluValLysLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142

Db 366 AGCACTGAAGTACACAGGAAGTCCATAAGCAAGCAATATGATGATGGGGAACATATATCA 425

Qy 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162

Db 426 AGAAATATTACTCTGATGGGTGAAAGGAGGGAATCTCTGCTTGGAGGATATATCCAGCT 485

Qy 163 IlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182

Db 486 GAGTCTCTGGTTAAATTTATTGGATGGAGATCCCTCATCTGAGCTGTATATCTGAGCCA 545

Qy 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202

Db 546 CTGCAATCAAGAGTTACAAAGTCCCAAGCACCATGAGGTCAATGCTCCGGTGAAGAA 605

```

Qy 203 GlyAspGlyGlyAsnAlaLysGlyArgLys-----LysAlaSerAsp 216
Db 606 ATCAATGAAGCCAGAAAGACGCTAGAAAGGTTGGTGAGATTATGGGTTTCATCTTCAAAA 665
Qy 217 AspValThrAspProArgValThrAsp----- 225
Db 666 GTTGTAATCTGCAAAAAGTACGATGATGCTGCTGCTGATATTGTTACTACATCTGGCCCT 725

Qy 226 -----ProProValAspAspGluArgLysAla 235
Db 726 GCTCTTCAACCAAGTTGATGTCACGACGCAAAAATCA 761

RESULT 3
CL795819 764 bp DNA linear GSS 06-AUG-2004
DEFINITION OR CBa0005M04.r OR CBa Oryza rufipogon genomic clone OR_CBa0005M04
3', genomic survey sequence.
ACCESSION CL795819
VERSION CL795819.1 GI:51017835
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 764)
AUTHORS Kim,H., Yu,Y., Wisotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
Jettly,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0005 row: M column: 04
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES
source
1..764
Location/Qualifiers
1..764
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR CBa0005M04"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/notes="vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII;
drk treated 36 hrs before harvest"

ORIGIN
Alignment Scores:
Pred. No.: 5.95e-24 Length: 764
Score: 317.50 Matches: 84
Percent Similarity: 50.00% Conservative: 42
Best Local Similarity: 33.33% Mismatches: 107
Query Match: 9.87% Indels: 19
DB: 9 Gaps: 2

US-10-018-929c-3_COPY_650_1270 (1-621) x CL795819 (1-764)

Qy 3 AspaAheLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 6 GATGCCATCATATATATGTTCTTGACACACACCACTTAACGACTTGAAGGCTCTTCAA 65

```

```

Qy 23 LysAlaLysLeuGluSerCysSerGluArgThrLysIlePheArgLeuTySerValCys 42
Db 66 AAGATCAAGTAGAGTACCAATTTGACGCTGTGAGCAATTTTCGCTTTGTACACCTTTT 125
Qy 43 ThrValGluGluLysAlaLeuLysLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 126 ACAGTTGAGGAGAAGAGTCTTGTCTTTCGAGGCGGCGCATTTGTTATTGATAACAACATA 185
Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyLeuPhe 82
Db 186 CAAGATCTAAGGACTAGTCTTAAACAATCTCTACTCAGATGGGCTGCTGGATGTGTGTT 245
Qy 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
Db 246 AGTAGACTTATGATGAGTCCAGCAAGATGACCAAGTCAAGTAAAGTTCTGAAATCGAGAGA 305
Qy 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlyGlu 122
Db 306 CACTTTATTGATGAGTAAATCGTTGAGTTCTTAAACAAGTTATCCCAACTGTTGAAGAC 365
Qy 123 GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTySer 142
Db 366 AGCACTGAAGTACACAGGAAGTCCATAAGCAAGCTAATATGAGTGGGAAGTATATATCA 425
Qy 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
Db 426 AGAATAATTAATCTGATGGGTGAAAGAGGAGGAATCTCTGTGCTTGAGGATAATCCAGCT 485
Qy 163 IlePheTrpSerLysLeuGluGlyLysAsnProMetTrpLysTyProSerAspThr 182
Db 486 GAGTTCTGGTAAATTTATTTGGATGGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 545
Qy 183 ProGlnArgAsnArgLysArgValGlnTyPheGluGlySerGluAlaSerProLysThr 202
Db 546 CTGCAATCAAGAGTTACAAAGTCCAGACCATGATGATGATGATGATGATGATGATGATG 605
Qy 203 GlyAspGlyGlyAsnAlaLysLysArgLys----- 212
Db 606 ATCAATGAAGCCAGAAAGCGTAGAAAGGTTGGTGAGATTATGGGTTTCATCTTCAAAA 665
Qy 213 -----LysAlaSerAspValThrAspProArgValThrAsp----- 225
Db 666 GTTGATCTGACAAAAGTAAAGATGATGATGATGATGATGATGATGATGATGATGATG 725
Qy 226 -----ProProValAspAspGluArgLysAla 235
Db 726 GCTCTTCAACCAAGTTGATGTCACGACGCAAAAATCA 761

RESULT 4
BH419713 369 bp DNA linear GSS 12-DEC-2001
LOCUS BOG0869TF BOGO Brassica oleracea genomic clone BOG0869, genomic
DEFINITION survey sequence.
ACCESSION BH419713
VERSION BH419713.1 GI:17605441
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 369)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOG0869TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.

```



```

mRNA sequence.
BG833433
VERSION BG833433.1 GI:14191203
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 951008 row: C column: 07.
FEATURES
    source
        1..525
            /organism="Zea mays"
            /mol_type="mRNA"
            /cultivar="BMS (Black Mexican Sweet)"
            /db_xref="taxon:4577"
            /tissue_type="suspension culture"
            /dev_stage="mixed logarithmic and stationary growth
            phases"
            /lab_host="DH10B"
            /clone_lib="951 - BMS tissue from Walbot Lab (GR)"
            /notes="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
            library was prepared by George Rudenko using poly (A)
            selected RNA and Universal Riboclone cDNA Synthesis System
            (Promega). cDNA was synthesized using both random and
            oligo(dT) primers in separate reactions and equipped with
            EcoRI adaptors. Library was size-fractionated on agarose
            gels (for insert size >400bp) and non-directionally cloned
            into EcoRI-digested pUC19 vector. Blue/white selection on
            carbenicillin-containing plates was used to recover
            positive clones."
ORIGIN
Alignment Scores:
Pred. No.: 3.27e-14 Length: 525
Score: 227.00 Matches: 53
Percent Similarity: 63.87% Conservative: 23
Best Local Similarity: 44.54% Mismatches: 37
Query Match: 7.06% Indels: 6
DB: 4 Gaps: 2
US-10-018-929C-3_COPY_650_1270 (1-621) x BG833433 (1-525)
Qy 3 AspalatheleuPheGlySerSerLeuAnProSerHisAspValLysHisValGlu 22
Db 94 GATGCCATCATATATTGTAGTACTGGGAATCCCATGAATGACTTGGAGGCCCTTCAG 153
Qy 23 LysileuLysleuGluSerCysSerGluArgThrLysilePheArgLeuTySerValCys 42
Db 154 AGGTCAGCATGGAGTCACATGCTGAACCTGTGCTATTTTCGTTTGTACTGCTCTTTT 213
Qy 43 ThrValGluGluLysAlaLeuLeuAlaArgGlnAnMetArgGlnAnLysAlaVal 62
Db 214 ACATGGAGGAAAAGCTCTATTCTTTCGAAGCATGACCATATTTGGATAGTAATATT 273
Qy 63 GluAnLeuAnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyLeuPhe 82
Db 274 GTGAATATACCCCTAGCTGAGCATTTGTCTGCTCAGTTGGGGCGCATCTTCTCTTC 333
Qy 83 AspLysLeuAspHisPhe-----HisSerGluThrProAspSerGlyValSerPhe 100
    ::::|||||

```

```

334 AATAGACTTGAGGAGTTGCAGCACATAGCTAC-----TCAATGCTCTGCT 381
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLys 119
    ::::|||||
Db 382 GATGAGCTTTTCATGGATAATGTCGATTGGAGTTTGGACAAATTTACTCAGCAAG 438
RESULT 7
BE203168
LOCUS EST403190 KV1 Medicago truncatula cDNA clone pKV1-4N3, mRNA
DEFINITION BE203168 569 bp linear EST 07-SEP-2000
sequence.
ACCESSION BE203168
VERSION BE203168.1 GI:8746439
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 569)
AUTHORS VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula 24 hours after inoculation
with Sinorhizobium meliloti
JOURNAL Unpublished (1999)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@csb.umn.edu
TIGR sequence name:MTIAK74TK
More information is available at:
http://chrwise.tamu.edu/medicago
Seq primer: Sknod (CTA GAA CTA gtg gat CC) .
FEATURES
    source
        1..569
            /organism="Medicago truncatula"
            /mol_type="mRNA"
            /cultivar="genotype A17"
            /db_xref="taxon:3880"
            /clone="pKV1-4N3"
            /tissue_type="Seedling roots"
            /dev_stage="24 hours post-inoculation with Sinorhizobium
            meliloti"
            /lab_host="E. coli strain XL0LR"
            /clone_lib="KV1"
            /notes="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:
            XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
            was directionally ligated into the Unizap XR vector from
            extracts. Plasmids containing cDNA inserts were excised
            from the recombinant lambda-zap phage using Ex-assist
            helper phage and propagated in XL0LR cells."
ORIGIN
Alignment Scores:
Pred. No.: 7.05e-14 Length: 569
Score: 224.50 Matches: 64
Percent Similarity: 42.27% Conservative: 29
Best Local Similarity: 29.09% Mismatches: 78
Query Match: 6.98% Indels: 49
DB: 2 Gaps: 6
US-10-018-929C-3_COPY_650_1270 (1-621) x BE203168 (1-569)
Qy 206 GlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgValThrAsp 225
Db 40 GGCACACAAATAGCGCTTGAAGTAAGCAACAAT-----

```

```
QY 226 ProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGlu 245
Db 76 ----AATGTGGATCAGCTTCATCAAAATCTGTAGTGAAAGATTACCAATGGAATCAAG 132
QY 246 SerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeu 265
Db 133 GCAGAT-----AAACCTCAGGGTAATGATGATTGAA 165
QY 266 AspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGlu 285
Db 166 TCCGGGAAAAAC----- 177
QY 286 AspMetLeuAlaSerGlnAspTTPGlyLysIleProAspGluSerGlnArgArgLeuHis 305
Db 178 -----GGCAGAGTCATCATGAG---CTGAGGAGCTTACAT 210
QY 306 ThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAlaCys 325
Db 211 CTTCGCTGAAGCCACAGATCACAGAGCTCTGTGAAGTCTCTCTTCCAGATAATGTT 270
QY 326 ThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu 345
Db 271 AGCAATGTGCTTGACAAATTTCTTGAGTATGTTACGAAACAAGCATCATTTCAATTGGGAA 330
QY 346 ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuVal 365
Db 331 ---TCTGTTTATACAGTCTTTTCAAAATATCTCTGATCAGGCTGCTCTCTTCTGCTA 387
QY 366 LysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLys 385
Db 388 AATCACAAGTTGACCAGAGACTATCTTGCTCTTTCAAAAGAACAAATTGAATTTTGAC 447
QY 386 CysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPhe 405
Db 448 TGTGAAGCAAGCAGAGGTGGACTATAGTTATTTCTTGCGAGTGTGTTGAAGAAATCTTT 507
QY 406 LeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerVal 425
Db 508 GTATATCGTACA-----GGCATTATTAATGATTAATTTCTTCTTGAAAGTT 552
```

```
RESULT 8
LOCUS BG833434 482 bp mRNA linear EST 23-MAY-2001
DEFINITION 95108C07.x3 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA, mRNA sequence.
```

```
ACCESSION BG833434
VERSION BG833434.1 GI:14191204
KEYWORDS EST.
```

```
SOURCE Zea mays
ORGANISM Zea mays
```

```
REFERENCE 1 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
TITLE clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 482)
Walbot V.
```

```
Maize ESTs from various cDNA libraries sequenced at Stanford University
```

```
Unpublished (1999)
```

```
Contact: Walbot V
```

```
Department of Biological Sciences
```

```
Stanford University
```

```
855 California Ave, Palo Alto, CA 94304, USA
```

```
Tel: 650 723 2227
```

```
Fax: 650 725 8221
```

```
Email: walbot@stanford.edu
```

```
Plate: 951008 row: C column: 07.
```

```
Location/Qualifiers
```

```
1. .482
```

```
/organism="Zea mays"
```

```
/mol_type="mRNA"
```

```
/cultivar="BMS (Black Mexican Sweet)"
```

```
/db_xref="taxon:4577"
```

```
FEATURES
```

```
source
```

```
/tissue type="suspension culture"
/dev stage="mixed logarithmic and stationary growth
phases"
/lab host="DH10B"
/clone lib="951 - BMS tissue from Walbot Lab (GR)"
/notes="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."
```

ORIGIN

```
Alignment Scores:
Pred. No.: 1.69e-13 Length: 482
Score: 220.00 Matches: 50
Percent Similarity: 64.29% Conservative: 22
Best Local Similarity: 44.64% Mismatches: 34
Query Match: 6.84% Indels: 6
DB: 4 Gaps: 2
```

```
US-10-018-929C-3_COPY_650_1270 (1-621) x BG833434 (1-482)
```

```
QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
```

```
Db 142 GATGCCATCATCATATATTTGTAGTACTGGAATCCCATGAATGACTTTGAGAGCCCTTCAG 201
```

```
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
```

```
Db 202 AGGTCAGCATGAGTCACAGCTCTGACCTGCTATTTTTCGTTGTTACTGCTCTTT 261
```

```
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
```

```
Db 262 ACAGTGGAGGAAAAGCTCTTATTTCTTCAAGCATGACCATATTTGGATAGTAATATT 321
```

```
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
```

```
Db 322 GTGAATATATAACCCCTGAGCCCATGTCTGCTCAGTTGGGCGCATCATTTCTCTTC 381
```

```
QY 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
```

```
Db 382 AATAGACTTGAGAGTTGACGACCAACATAGCTAC-----TCAATGCTCTCTGGT 429
```

```
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
```

```
Db 430 GATGAGCTTTTCATGATAATGTAGATTGAGGATTT 465
```

RESULT 9

```
BP133890
```

```
LOCUS BP133890
```

```
DEFINITION BP133890 MAT001 Nicotiana tabacum cDNA clone BY6204, mRNA sequence.
```

```
ACCESSION BP133890
```

```
VERSION BP133890.1 GI:32876775
```

```
KEYWORDS EST.
```

```
SOURCE Nicotiana tabacum (common tobacco)
```

```
ORGANISM Nicotiana tabacum
```

```
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
asterids; lamids; Solanales; Solanaceae; Nicotiana.
```

```
1 (bases 1 to 668)
```

```
Matsuoka,K., Tashiro,G., Horiguchi,T., Demura,T. and Fukuda,H.
```

```
Profiling growth-phase dependent gene expression of tobacco BY-2
```

```
cells by comprehensive microarray analysis
```

```
Unpublished (2003)
```

```
Contact: Ken Matsuoka
```

```
Morphogenesis Research Group
```

```
RIKEN Plant Science Center
```

```
1-7-2 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
```

```
Tel: 81-45-503-9575
```



```

JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT      Contact: Taku Demura
              Morphogenesis Research Group
              RIKEN Plant Science Center
              1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
              Tel: 81-45-503-9605
              Fax: 81-45-503-9573
              Email: demura@postman.riken.go.jp
              This clone was obtained at our laboratory.
              Seq primer: M13 forward.

FEATURES     source
              1..516
               /organism="Zinnia elegans"
               /mol_type="mRNA"
               /cultivar="Canary bird"
               /db_xref="taxon:34245"
               /clone="Z2904"
               /tissue_type="mesophyll cell"
               /clone_lib="zinnia cultured mesophyll cell equalized cDNA"
               /note="vector: pGEM-T easy; cultured in tracheary element
               differentiation-inductive medium"

ORIGIN
Alignment Scores:
Pred. No.:      1.64e-12      Length:      516
Score:          211.50      Matches:    51
Percent Similarity: 58.33%      Conservative: 19
Best Local Similarity: 42.50%      Mismatches: 46
Query Match:     6.58%      Indels:     4
DB:              1          Gaps:           1

US-10-018-929c-3_COPY_650_1270 (1-621) x AU288479 (1-516)
Qy 299 GlsrGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnVal 318
    |||
    |||
Db 82 GAAGGGAAATGAACCTTTAAGGACCTCTGAAGCTTAAATAACCTAACTATGTGAAGTC 141
    |||
Qy 319 LeuHisLeuSerArgAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
    |||
Db 142 CTTAAGTTGTCGGAGGATGGAAGATAATGTTGGAAGGATTTCTCGAGTATGTTATTGAG 201
    |||
Qy 339 AsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSer 358
    |||
Db 202 AATTATCGCTCACAGGAACCTGAAAGCATATACATGCTTCTTGATATCCCTGTGT 261
    |||
Qy 359 -TrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAl 376
    |||
Db 262 TNGGATTGGTTCGGGATTTGTAACACACAGCTTGATAGGAAGCAATCCATAAGTCTTGC 321
    |||
Qy 378 aAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLe 398
    |||
Db 322 GAACGAACACTTGGGTTTCAGTTGTAAGGAATAATGAAGCAACTCAGTGTACTCGAAGCT 381
    |||
Qy 398 uSerCysMetLysSerLeuPheLeuGluHisThr-----GlnGlyLeuGlnPhe 414
    |||
Db 382 GGAATTTCGAAGAGATGTTCTTACTTCACACGGACATCCAAAGAACCTTCGTTTC 439
    |||

RESULT 12
A1947519
LOCUS      603022G06.x1 603 - stressed root cDNA library from Wang/Bohnert lab
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION A1947519.1 GI:5739829
VERSION    A1947519
KEYWORDS   EST.
SOURCE     Zea mays
           Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 486)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL      University
COMMENT      Unpublished (1999)
              Contact: Walbot,V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Plate: 603022 row: G column: 06.

FEATURES     source
              1..486
               /organism="Zea mays"
               /mol_type="mRNA"
               /cultivar="B73"
               /db_xref="taxon:4577"
               /tissue_type="seedling"
               /dev_stage="salt stress"
               /lab_host="E. coli XL Gold"
               /clone_lib="603 - stressed root cDNA library from
               Wang/Bohnert lab"
               /note="Organ: root; Vector: pBluescriptII SK(+/-) XR;
               Seedling stressed root cDNA library from Wang/Bohnert lab"

ORIGIN
Alignment Scores:
Pred. No.:      4.13e-12      Length:      486
Score:          207.50      Matches:    56
Percent Similarity: 53.21%      Conservative: 27
Best Local Similarity: 35.90%      Mismatches: 64
Query Match:     6.45%      Indels:     9
DB:              1          Gaps:           4

US-10-018-929c-3_COPY_650_1270 (1-621) x A1947519 (1-486)
Qy 44 ValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGlu 63
    |||
Db 1  GTGGAGGAAAAGCTCTTATCTTGCAAGCATGACCATATTTGGATAGTAAATATTGTG 60
    |||
Qy 64 AsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAsp 83
    |||
Db 61 AATATAACCCCTAGCCTGAGCCATTGTGCTCAGTTGGGGCGCATCATTTCTTCCAAT 120
    |||
Qy 84 LysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPheGlu 101
    |||
Db 121 AGACTTGAGGAGTTGCGACGCAACATAGCTAC-----TCAATGTCCTGTGTGAT 168
    |||
Qy 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
    |||
Db 169 GAGCTTTTCATGATAATGTAGATTGGAGTTTTCACAAAATTTACTCAGCAAGGTTGAA 228
    |||
Qy 122 GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
    |||
Db 229 CTCAGAACTGAATCAGGCAATACAGCCATATCAAGCTTATCTGTGTGGATCTTTCTAT 288
    |||
Qy 142 SerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGluGluSer 160
    |||
Db 289 TCTAGGGCTATTGTTAGCGGTGAGAGGAGGAATTCATCAGTTGATGTGACTG 348
    |||
Qy 161 ProAsnIle-----PheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyr 178
    |||
Db 349 CCAAAGTCTGGGCCCTACTGTTAAGTTTACTTAAATGGGAGGCTCTCTCAGTGGCAGTAT 408
    |||
Qy 179 ProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
    |||
Db 409 ATATCTGAGCCAGTACAAAGAGCCGTAAGAAGATAATAAACAATGGAA 456
    |||

RESULT 13
B21272
LOCUS      B21272
DEFINITION F3G11-T7.1 IGF Arabidopsis thaliana genomic clone F3G11, genomic
            survey sequence.
ACCESSION B21272

```

```

VERSION B21272.1 GI:2396326
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 799)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: F3G11-T7, F3G11-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 82
High quality sequence stop: 116.
FEATURES
    source
        1. 799
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clones="F3G11"
            /sex="hermaphrodite"
            /clone_lib="IGP"
            /note="Vector: BelohACII; Site_1: EcoRI; Site_2: EcoRI;
            produced by Thomas Altmann"
ORIGIN
Alignment Scores:
Pred. No.: 6.28e-11 Length: 799
Score: 200.00 Matches: 74
Percent Similarity: 58.99% Conservative: 8
Best Local Similarity: 53.24% Mismatches: 55
Query Match: 6.22% Indels: 7
DB: 8 Gaps: 1
US-10-018-929C-3_COPY_650_1270 (1-621) x B21272 (1-799)
Qy 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
Db 375 CGAGCTGATGGTTATCTCTGTT-GGGAGCAGCTAGAAATCCATCGCATGATGTTAGCAC 433
Qy 21 Val-GluLysIleGlySerGlySerGluArgThrLysIlePheArgLeuTyrSe 40
Db 434 GTAAAGAGNANNANNTCCAGTCTCTTTCTGGTACTATGATATCTATGTTGTC 493
Qy 40 rValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLy 60
Db 494 AGTATGTTTCAGTGAAGAAAGCGGTGATCTGGCTAGGCANNATAG-CGGCAANTNA 552
Qy 60 sAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTy 80
Db 553 GGCTGT-GAAGACCTANACCGCTCTCTCAGCAGCAGCTGCTCATGTTGGGGCTTCATA 611
Qy 80 rLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPh 100
Db 612 CTNATTTGATNAACTGTTCTTTTCCACNCANNTGAATCCCAA--TCCGGAAGTCCCAT 669
Qy 100 eGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysG 120
Db 670 TGANCATCTTNATGATGAACGGCGTATCCT-GANTCCTCTGTCATCATCTCTCTCCCAAGG 728
Qy 120 yGlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGln 138

```

```

Db 729 TTAANANNATGAT---TCAANCTGTGTCTCTCTTTGGAGGACNACCATGCTCAG 780
RESULT 14
BH788688/c 560 bp DNA linear GSS 28-MAR-2002
LOCUS fmb021f001a02k0 fmb filtered library Zea mays genomic clone
DEFINITION fmb021f001a02 5', genomic survey sequence.
ACCESSION BH788688
VERSION fmb021f001a02
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 560)
AUTHORS Budiman,M.A., Freese,R.G., Bedell,J.A., Nunberg,A.N. and Lakey,N.D.
TITLE Genethresher methylation filtered genomic sequences from maize
JOURNAL Unpublished (2002)
COMMENT Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: fmb021f001 row: e column: 02
Seq primer: SK reverse
Class: shotgun
High quality sequence stop: 560.
FEATURES
    source
        1. 560
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /cultivar="Mol7"
            /db_xref="taxon:4577"
            /clone="fmb021f001a02"
            /clone_lib="fmb filtered library"
            /note="Organ: leaf; Vector: pBSK(-); Site_1: HincII; DNA
            prepared from purified nuclei was randomly sheared,
            end-repaired, size fractionated to enrich for the 0.5 to
            5 kb fraction, ligated into HincII-digested pBSK(-)
            vector and electroporated into E. coli cells."
ORIGIN
Alignment Scores:
Pred. No.: 5.84e-11 Length: 560
Score: 198.00 Matches: 52
Percent Similarity: 61.98% Conservative: 23
Best Local Similarity: 42.98% Mismatches: 38
Query Match: 6.16% Indels: 8
DB: 8 Gaps: 2
US-10-018-929C-3_COPY_650_1270 (1-621) x BH788688 (1-560)
Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 379 GATGCCATCATATATTTAGTACTGGAATCCCATGAATGACTTGGAGGCCCTTCAG 320
Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 319 AGGGTCAGCATGGAGTCACAGTCTGAACCTGTGCCTATTTTTCGTTGTACTCGCTTTT 260
Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsn-MetArgGlnAsnLysAlaVa 62
Db 259 ACAGTGAGGAAAGAGCTTTATTTTCGACGATGACCATTTTGGATNAGTATAT 200
Qy 62 lGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMet-TrpGlyAlaSerTyrLeuP 82
Db 199 TGTGAATATAACCCCTAGCTGAGCCATTTCTTGGCCAGTGTGGGGCGCATCATTTCTCT 140
Qy 82 heAspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerP 100

```

```
Db      139 TCAATAGACTTGAGGAGTTGCAGCAACATAGCTAC-----TCAATAGTCTCTG 92
Qy      100 heGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIys 119
Db      91 GTGATGAGCTTTTCATGGATTAATGTAGATTGGAGTTTTTGGACAAAATTACTCAGCAAG 33

RESULT 15
CA215647
LOCUS   CA215647
DEFINITION CA215647 669 bp mRNA linear EST 25-SEP-2003
          SCRLAD1140B12.g AD1 Saccharum officinarum cDNA clone SCRLAD1140B12
          5', mRNA sequence.
ACCESSION CA215647
VERSION   CA215647.1 GI:35264132
KEYWORDS  EST.
SOURCE    Saccharum officinarum
ORGANISM  Saccharum officinarum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
          complex.
REFERENCE 1 (bases 1 to 669)
AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE     The libraries that made SUCEST
JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT   Contact: Arruda P
          Centro de Biologia Molecular e Engenhariaia Genetica
          Universidade Estadual de Campinas
          Caixa Postal 6010, 13083-970, Campinas SP, Brazil
          Tel: 55 19 3788 1137
          Fax: 55 19 3788 1089
          Email: parruda@unicamp.br
          Clone distribution: clone distribution information can be found
          through the Brazilian Clone Collection Center (BCCC) at
          http://www.bccccenter.fcav.unesp.br
          Plate: 140 row: B column: 12
          Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source            1..669.
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCRLAD1140B12"
                     /lab_host="DH10B"
                     /clone_lib="AD1"
                     /note="Organ: seedlings inoculated with Gluconacetobacter
                     diazotrophicans; Vector: pSport1; Site1: SalI; Site_2:
                     NotI; An unidirectional cDNA library generated from
                     [seedlings inoculated with Gluconacetobacter
                     diazotrophicans]. cDNA was prepared from polyA+ mRNA using
                     Superscript Plasmid System Kit (Invitrogen). The
                     double-strand cDNAs were fractionated in a sepharose
                     CL-2B 40cm-columns and fragments sizing between 0.8 and
                     1.5 Kb were directionally cloned into the vector. Details
                     of each source of RNA and library construction can be
                     obtained at http://sucest.lad.ic.unicamp.br/public"
```

```
Db      491 AAGCTCAAGTAGGGTCAACAACCTGAAGTATGTCCGCAATTTCCGCTTATATATCTCCTTC 550
Qy      43 ThrValGluGluLysAlaLeuIleLeuAlaAArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db      551 ACAGTGGAGGAGAGAGTCTTGTGCTCGCCCAAGCAAGCATGATTATTGTATAAGCACAGC 610
Qy      63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db      611 CAAGACATAACATCTAGTTTAAGCCATTGCTTGGTCAGCTGGGGGTGTATCATCTCCTC 667

Search completed: July 11, 2005, 23:52:22
Job time : 5390.33 secs
```

```
Db      139 TCAATAGACTTGAGGAGTTGCAGCAACATAGCTAC-----TCAATAGTCTCTG 92
Qy      100 heGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIys 119
Db      91 GTGATGAGCTTTTCATGGATTAATGTAGATTGGAGTTTTTGGACAAAATTACTCAGCAAG 33

RESULT 15
CA215647
LOCUS   CA215647
DEFINITION CA215647 669 bp mRNA linear EST 25-SEP-2003
          SCRLAD1140B12.g AD1 Saccharum officinarum cDNA clone SCRLAD1140B12
          5', mRNA sequence.
ACCESSION CA215647
VERSION   CA215647.1 GI:35264132
KEYWORDS  EST.
SOURCE    Saccharum officinarum
ORGANISM  Saccharum officinarum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
          complex.
REFERENCE 1 (bases 1 to 669)
AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE     The libraries that made SUCEST
JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT   Contact: Arruda P
          Centro de Biologia Molecular e Engenhariaia Genetica
          Universidade Estadual de Campinas
          Caixa Postal 6010, 13083-970, Campinas SP, Brazil
          Tel: 55 19 3788 1137
          Fax: 55 19 3788 1089
          Email: parruda@unicamp.br
          Clone distribution: clone distribution information can be found
          through the Brazilian Clone Collection Center (BCCC) at
          http://www.bccccenter.fcav.unesp.br
          Plate: 140 row: B column: 12
          Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source            1..669.
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCRLAD1140B12"
                     /lab_host="DH10B"
                     /clone_lib="AD1"
                     /note="Organ: seedlings inoculated with Gluconacetobacter
                     diazotrophicans; Vector: pSport1; Site1: SalI; Site_2:
                     NotI; An unidirectional cDNA library generated from
                     [seedlings inoculated with Gluconacetobacter
                     diazotrophicans]. cDNA was prepared from polyA+ mRNA using
                     Superscript Plasmid System Kit (Invitrogen). The
                     double-strand cDNAs were fractionated in a sepharose
                     CL-2B 40cm-columns and fragments sizing between 0.8 and
                     1.5 Kb were directionally cloned into the vector. Details
                     of each source of RNA and library construction can be
                     obtained at http://sucest.lad.ic.unicamp.br/public"
```

```
ORIGIN
Alignment Scores:
Pred. No.:      3.6e-10      Length:      669
Score:          192.00      Matches:    37
Percent Similarity: 73.42%      Conservative: 21
Best Local Similarity: 46.84%      Mismatches: 21
Query Match:    5.97%      Indels:      0
DB:              6          Gaps:          0

US-10-018-929C-3_COPY_650_1270 (1-621) x CA215647 (1-669)

Qy      3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisPheValLysHisValGlu 22
Db      431 GATTCATTAATCATTAATGAGTAGTGAATCCACTAATGAATGATCTGAAGGCTCTTCGA 490
Qy      23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 19:44:24 ; Search time 61.6667 Seconds
(without alignments)
5156.776 Million cell updates/sec

Title: US-10-018-929C-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILFGSLNPSHDVHKH.....SGNNKHGSGSNISQNAPDV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3216	100.0	2001	2 Q9M659	Q9M659 arabidopsis
2	3207	99.7	2001	2 Q9M658	Q9M658 arabidopsis
3	3207	99.7	2254	2 Q9LNO2	Q9LNO2 arabidopsis
4	163.5	5.1	1461	2 O17909	O17909 caenorhabdi
5	162.5	5.1	842	2 O6ZU43	O6ZU43 homo sapien
6	162.5	5.1	1739	1 CHD2_HUMAN	CHD2_HUMAN
7	160.5	5.0	1268	2 Q7RG05	Q7RG05 plasmodium
8	160	5.0	1738	2 Q9SI41	Q9SI41 arabidopsis
9	158	4.9	4007	2 Q7SH24	Q7SH24 neurospora
10	157	4.9	1245	2 O8IAN1	O8IAN1 plasmodium
11	156	4.9	1320	2 Q9JK25	Q9JK25 rattus norv
12	155.5	4.8	2230	1 G0A4_HUMAN	G0A4_HUMAN
13	154	4.8	905	2 Q6BHF8	Q6BHF8 debaryomyce
14	154	4.8	2757	2 Q7RRR9	Q7RRR9 plasmodium
15	152.5	4.7	1330	2 Q22342	Q22342 caenorhabdi
16	152	4.7	2067	2 Q9U3U8	Q9U3U8 plasmodium
17	152	4.7	2160	2 Q8IE50	Q8IE50 plasmodium
18	151.5	4.7	1310	2 Q949K0	Q949K0 lycopersico
19	151	4.7	1611	2 Q7RD43	Q7RD43 plasmodium
20	151	4.7	7210	2 Q9V7G8	Q9V7G8 drosophila
21	151	4.7	9270	2 Q8MLD9	Q8MLD9 drosophila
22	150.5	4.7	1033	2 Q676A5	Q676A5 oikopleura
23	150	4.7	979	2 Q9EP71	Q9EP71 mus musculu
24	150	4.7	1500	2 Q75J40	Q75J40 oryza sativ
25	150	4.7	1786	2 Q7ZT34	Q7ZT34 poephila gu
26	149.5	4.6	992	2 Q9NTH6	Q9NTH6 homo sapien
27	148	4.6	1456	2 Q9UPV0	Q9UPV0 homo sapien
28	147.5	4.6	1927	2 Q75WV9	Q75WV9 helicobacte
29	147	4.6	962	2 Q96164	Q96164 plasmodium
30	147	4.6	1939	1 MYH4_HUMAN	MYH4_HUMAN
31	147	4.6	2042	2 Q6BUQ9	Q6BUQ9 debaryomyce

RESULT 1

Q9M659 PRELIMINARY; PRT; 2001 AA.

AC Q9M659; 628 2 Q6CAJ4

DT 01-OCT-2000 (TrEMBLrel. 15, Created) 751 2 Q8JRV5

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update) 1407 2 Q7QCA4

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update) 145.5 4.5 1931 2 Q9NCF9

DE MOM. 145.5 4.5 1931 2 Q9VKN9

OS Arabidopsis thaliana (Mouse-ear cress). 145.5 4.5 6761 2 Q6LEX8

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; 145 4.5 738 2 Q6CDT3

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; 145 4.5 1755 2 Q6CPF6

OC euroids II; Brassicales; Brassicaceae; Arabidopsi. 145 4.5 1960 1 MYH9_HUMAN

OX NCBI_TaxID=3702; 145 4.5 3187 2 Q63714

RP SEQUENCE FROM N.A. 145 4.5 8407 2 Q7RTM4

RX MEDLINE=20279299; PubMed=10821279; DOI=10.1038/35012108; 145 4.5 743 2 Q8WP33

RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.; 144.5 4.5 1938 2 Q9PVE1

RT "Disruption of the plant gene MOM releases transcriptional silencing 144 4.5 887 2 Q75AF5

RL Nature 405:203-206(2000).

DR EMBL; AF213627; AAF73380.1; .

SQ SEQUENCE 2001 AA; 218585 MW; F6C5DAE897F2FE61 CRC64;

Query Match 100.0%; Score 3216; DB 2; Length 2001;
Best Local Similarity 100.0%; Pred. No. 5.4e-171;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RADAFILFGSLNPSHDVHKHVKIKIESCSERTKIFRLYSVCTVEEKALILARONRK 60
Db 650 RADAFILFGSLNPSHDVHKHVKIKIESCSERTKIFRLYSVCTVEEKALILARONRK 709

Qy 61 AVENLARSLTHALLMWCASYLFDKLDHFHSSSTPDSGVSEFQSIMDGVHFEFSSILSKG 120
Db 710 AVENLARSLTHALLMWCASYLFDKLDHFHSSSTPDSGVSEFQSIMDGVHFEFSSILSKG 769

Qy 121 GEENEVLCILLLEAKHAQGYTSSDSTLFGDHIKLSDEESPNI FWSKLGGNPMWKYPS 180
Db 770 GEENEVLCILLLEAKHAQGYTSSDSTLFGDHIKLSDEESPNI FWSKLGGNPMWKYPS 829

Qy 181 DTPQRNRKRYQVEGSEASPKTGGGNKKRKKASDDVTDPRVTDPPVDDDERKASGDH 240
Db 830 DTPQRNRKRYQVEGSEASPKTGGGNKKRKKASDDVTDPRVTDPPVDDDERKASGDH 889

Qy 241 MGALESFKVITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWKIPDES 300
Db 890 MGALESFKVITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWKIPDES 949

Qy 301 QRRHLTVLKPMAKLCQVLHLSDACTSMVGNFLEYIENHRIYEEPTATTFOAFOIALSWI 360
Db 950 QRRHLTVLKPMAKLCQVLHLSDACTSMVGNFLEYIENHRIYEEPTATTFOAFOIALSWI 1009

Qy 361 AALLVKQILSHKSLVPSANSEAFKCSRVVDYTYTSLSCMKSLFLEHTQGLQDCGCTN 420
Db 1010 AALLVKQILSHKSLVPSANSEAFKCSRVVDYTYTSLSCMKSLFLEHTQGLQDCGCTN 1069

ALIGNMENTS

```
QY 421 SKQSVVSTKLWNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 480
Db 1070 SKQSVVSTKLWNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 1129
QY 481 GIKKKYKQVQKLVQVHEEKKMELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 540
Db 1130 GIKKKYKQVQKLVQVHEEKKMELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 1189
QY 541 KLLDHNRYERFDEIKSEKNECLSKLEOMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 600
Db 1190 KLLDHNRYERFDEIKSEKNECLSKLEOMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 1249
QY 601 QSGNNKHFSGSSNISQNPADV 621
Db 1250 QSGNNKHFSGSSNISQNPADV 1270

RESULT 2
Q9M658 PRELIMINARY; PRT; 2001 AA.
AC Q9M658
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MOM.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopses.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20279299; PubMed=10821279; DOI=10.1038/35012108;
RA Amedeo P., Habu Y., Akbar K., Mittelsten Scheid O., Paszkowski J.;
RT "Disruption of the plant gene MOM releases transcriptional silencing
of methylated genes.";
RL Nature 405:203-206(2000).
DR EMBL; AF213628; AAF73381.1; -.
SQ SEQUENCE 2001 AA; 218567 MW; B9C85D4E62704441 CRC64;
```

```
Query Match 99.7%; Score 3207; DB 2; Length 2001;
Best Local Similarity 99.7%; Pred. No. 1.7e-170;
Matches 619; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RADAFILFGSSLNPSHDVHKVEKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 60
Db 650 RADAFILFGSSLNPSHDVHKVEKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 709
QY 61 AVENLNRSUTHALLMWGASYLFDKLDHFSSETPDGSGVSEQSIMDGVIEHFSILSSKG 120
Db 710 AVENLNRSUTHALLMWGASYLFDKLDHFSSETPDGSGVSEQSIMDGVIEHFSILSSKG 769
QY 121 GEENEVKLCILLAKHAQGYSSDSTLFGEDHIKLSDEESPNFWSKLLGKGNPMWKYPS 180
Db 770 GEENEVKLCILLAKHAQGYSSDSTLFGEDHIKLSDEESPNFWSKLLGKGNPMWKYPS 829
QY 181 DTQRNKRQVYEGESASPKTGDGNGAKRKASDDVTPRVTDPPVDDDERKASGDH 240
Db 830 DTQRNKRQVYEGESASPKTGDGNGAKRKASDDVTPRVTDPPVDDDERKASGDH 889
QY 241 MGALES PKVTLTQSSCKSGTDTLGDNDAFGLYSMGSHISGIPEDMLASQDNGKIPDES 300
Db 890 MGALES PKVTLTQSSCKSGTDTLGDNDAFGLYSMGSHISGIPEDMLASQDNGKIPDES 949
QY 301 QRLHVTLPKPMARKLCOVLHLSDACTSMVGNFLEYVIENHRIYEEPATTFQAQIALSWI 360
Db 950 QRLHVTLPKPMARKLCOVLHLSDACTSMVGNFLEYVIENHRIYEEPATTFQAQIALSWI 1009
QY 361 AALLVQILSHKESLVRANSELAPKCSRVEVDYIYSILSCWKSILFLEHTQGLQPCFGTN 420
Db 1010 AALLVQILSHKESLVRANSELAPKCSRVEVDYIYSILSCWKSILFLEHTQGLQPCFGTN 1069
```

```
QY 421 SKQSVVSTKLWNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 480
Db 1070 SKQSVVSTKLWNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 1129
QY 481 GIKKKYKQVQKLVQVHEEKKMELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 540
Db 1130 GIKKKYKQVQKLVQVHEEKKMELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 1189
QY 541 KLLDHNRYERFDEIKSEKNECLSKLEOMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 600
Db 1190 KLLDHNRYERFDEIKSEKNECLSKLEOMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 1249
QY 601 QSGNNKHFSGSSNISQNPADV 621
Db 1250 QSGNNKHFSGSSNISQNPADV 1270

RESULT 3
Q9LN02 PRELIMINARY; PRT; 2254 AA.
AC Q9LN02
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T6D22.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopses.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
Ecker J.R.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026875; AAF79839.1; -.
DR FIR; D86215; D86215.
DR HSP; P00918; 1B3.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_Coanhd.
DR InterPro; IPR011009; Kinase_like.
DR Pfam; PF00194; Carb_ahydase; 1.
DR ProDom; PD00865; Euk_Coanhd; 1.
SQ SEQUENCE 2254 AA; 248157 MW; 64E80CF274791FF3 CRC64;

Query Match 99.7%; Score 3207; DB 2; Length 2254;
Best Local Similarity 99.7%; Pred. No. 2e-170;
Matches 619; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RADAFILFGSSLNPSHDVHKVEKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 60
Db 883 RADAFILFGSSLNPSHDVHKVEKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 942
QY 61 AVENLNRSUTHALLMWGASYLFDKLDHFSSETPDGSGVSEQSIMDGVIEHFSILSSKG 120
Db 943 AVENLNRSUTHALLMWGASYLFDKLDHFSSETPDGSGVSEQSIMDGVIEHFSILSSKG 1002
```



```
RESULT 5
Q6ZU43 PRELIMINARY; PRT; 842 AA.
AC Q6ZU43
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ44007.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK125995; BAC86384.1; -.
SQ SEQUENCE 842 AA; 94360 MW; F195761EFDD8E33B CRC64;

Query Match 5.1%; Score 162.5; DB 2; Length 842;
Best Local Similarity 20.1%; Pred. No. 0.84;
Matches 123; Conservative 98; Mismatches 221; Indels 169; Gaps 26;

QY 122 EENEVKLCLEAKHAQGYSSDSTLFGEDHKLSDSEPNFWSKLLGG----- 171
DB 245 EKSEPKICRLNLTVPKADPT-----GSEPAKASEKEAPE---DTVDAGEESRREAA 293

QY 172 KNPWKY-----PSDTPQ-----RNRKRVQYFEGSEAPKTDGNGNAKRRKASDDVD 220
DB 294 KEPKKASALEEGSSDASQLEISEHMKEPQLSDSIASDPKSPHGLDFGFRSRISEHL 353

QY 221 PRVTD-----PVDDERKASGDHMGAL-----PKVITLQ 253
DB 354 VDVLSPVLGGACRAQOPLGIEDKDDQSSQDELQSKQSGLEERYHRLSPPLPHERAQ 413

QY 254 SSCKS-----SGTDGTL-----GNDAGFLYSGMSHIS-----IPDMLA--- 289
DB 414 SPRSLATEPEPQGPGEQPEWKEAEELGSDASLSQLQRRSTPEVPAPEQLESEA 473

QY 290 -----SDQMGKIPDESQRRLHTVLKPKM-----AKLCQ-----VLHL-----SDACTSMVGNFL 333
DB 474 LKAMEEAVAQVLEQDQRHLLSKQEKMQQLREKLQEBEEBILRLHQKQESLSLRL 533

QY 334 EYVIENHRIYEEBATTQAFQIALSWIAALLVKQLSHKESLVTRANSELAFKCSRVEDY 393
DB 534 QKAIE-----EEEARMEESQRLSWLRA--QVSSSTQADEQDQRAEQEASLQKLRELE- 586

QY 394 IYSILCMK-----SLFLEHTQGL-----QFDCFOTNSKQSVST-----KLVNESLSGA 438
DB 587 -----SQOKERASLEQKNQMLQLEKEEIEA-SEKSEQAALNAKEKALQQLREQLEGE 640

QY 439 TVRDEKINTKMRNSSDEECMTKRCRSHYSTARDI-----EKTISGKKYKQKQKLV 494
DB 641 --RKEAVATLEKHSAB-----LERLCSLSLEAKHREVVSLSLQKKIQEAAQKQKCL 693

QY 495 QEHEEK-----KMLLNAYADKKQLE-----TSKSVEA 523
DB 694 GQVEHRVHQSYHVAGYEHELSLSLREKQREVEGEHEHRRLLDKWKEHQVNAKAREQYEA 753

QY 524 AVTRITCSRTSTQVGDULKLDHNYRKFDEIKSEKNECLSKLEQMHEVAKKLAEDBACW 583
DB 754 EERKQRAELGLHTGELERLQRAHERLETVRQHQKLEDLRRRHRERQERKQLDLELD 813

QY 584 INRIKSWAKL 594
DB 814 ETRAKDVKARL 824

RESULT 6
CHD2_HUMAN STANDARD; PRT; 1739 AA.
AC Q14647;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein 2 (CHD-2).
GN Name=CHD2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97470991; PubMed=9326634; DOI=10.1073/pnas.94.21.11472;
RA Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT "Characterization of the CHD family of proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC -!- FUNCTION: Sequence-selective DNA-binding protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF006514; AAB87382.1; -.
DR HSSP; P23197; LAP0.
DR Genew; HGNC:1917; CHD2.
DR MIM; 602119; -.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR002464; DEAH box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; Chromo; 2.
DR Pfam; PF00176; SNF2 N; 1.
DR Pfam; PF00176; SNF2 N; 1.
DR PROSITE; PS00598; CHROMO 1; 2.
DR PROSITE; PS00013; CHROMO 2; 2.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW ATP-binding; DNA-binding; Helicase; Hydrolase; Nuclear protein;
KW Repeat.
FT DOMAIN 1 74 Ser-rich.
FT DOMAIN 121 139 Ser-rich.
FT DOMAIN 261 353 Chromo 1.
FT DOMAIN 378 456 Chromo 2.
FT NP_BIND 509 516 ATP (Potential).
FT SITE 617 620 DEAH box.
SQ SEQUENCE 1739 AA; 200560 MW; DG2089C35EBAA40F CRC64;

Query Match 5.1%; Score 162.5; DB 1; Length 1739;
Best Local Similarity 18.9%; Pred. No. 2.1;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

QY 2 ADAPFLGSLNPSHDVQVHEKIKIESCSERTKIFRLYSVCTVEKA-----L 49
DB 879 ADTVFVDSWNPQNDLQAAHRIQKQVNIYELVTGKTVEEIIERAKKKWLDHL 938

QY 50 ILARQNQRNKAIVEN-----LNRLTHALLMNGASYLFKLDHFSHSETPDGVSF 100
DB 939 VIQRMDTGTTRILENNSGRSNPNFKBELTALFKGAEDLPKELE-----GEES 988
```

Qy	101	EQSIMDVIHFFS	ILSSKGBENEVK	LCULLLAKAQTG	YSDDSTLFGEDH	KLSDSES	160					
Db	989	EPOEMD	---	IDEILFLAETRE	NVSTSATDELL	-SQFKVANFATMEDEEL	---EER 1038					
Qy	161	PNIFWSKLLG	KNPMWYPSDT	FORNKRQVFE	---	---	GSASPKTG 203					
Db	1039	PHKDWEII	---	---	PEQRKKVVEER	QKLEEIYMLPRISSTKKAQTND	S 1085					
Qy	204	DGNAKRRKKA	DDVTDPRVTD	PPVDDERKASG	KOHMGALSPKVIT	LQSSCKSGTGD	263					
Db	1086	DSDTESKROA	RSASASETED	---	SDDDKX	---	KRRGRPRSVRKOLVEG 1130					
Qy	264	TLDGN	---	DAFLYSMGSH	ISGIPEDML	---	ASDWGKIPDESQRRLHTVLKPKWAKLC 316					
Db	1131	FTDAEIRRF	TAKYKF	---	GLPLURECTIARD	-AELVDKSVADL	-----KRLG 1174					
Qy	317	QVLHLSDACT	SMVGNFLEY	VIEHNRIYEE	PATTFQAFIAL	SALMALLKQIILSHKESLV	376					
Db	1175	ELIHL	---	NSCVSAQME	VEEQLEKENSEG	KPGKR-RGPTIKISGV	-QVNWKSIIQHHEEPE 1230					
Qy	377	RANSEL	---	AFKCSRVEVD	YIYSILSCMKS	SLFLEHTQGLQFCFGTNSK	422					
Db	1231	MLHKSIPVD	PEEKKYCLT	CRKAAHFDV	EWGVEDDSRL	LLIGIYERHYG	-NWELIKTDP 1289					
Qy	423	QSVVSYKL	VNESLSGAT	VRDEKINTK	SMRNSSE	---	---	DEECMTEKR 464				
Db	1290	---	LKLTDLIL	PVET	---	DKPOGKOLQ	TRADYLLKLRLKLEKKGAVTGSSEAKLKR 1342					
Qy	465	---	---	---	CSHYSTATRDI	EKTISGIKKKYKQVOKLQVHEHEKK	501					
Db	1343	KPVAKKEN	KYPRLKEHG	IELSPRISDN	PSSEGEVKDGL	ESKSPMKKKQK-KKENKENK	1401					
Qy	502	MELLNMYADK	---	---	KOKLETSGVBA	---	---	AVIRITCSRTSQVGD	539			
Db	1402	EKQMSRKDK	EGDKERKKS	KDKKEKPSG	DAKSSSKRSQGP	VHITAGSEPVPIGEDED	1461					
Qy	540	---	---	---	LKLLDH	---	---	---	---	EIKSEKNECLKS	564	
Db	1462	DDLQDFTFS	ICKERM	PRVKALKOLQ	KDPKGLNVQ	EQLEHTRNCLIKIGDRIAECLKAYS	1521					
Qy	565	---	---	---	---	LEOMHEVAKKGLAEDEACWINR	IKSWAALKVCV	598				
Db	1522	QDEHILWR	NRNLWIFV	SKPTFED	PARKLHLYKWAHKKR	SGEEZ	-----	---	---	EQKKDDV	1572	
Qy	599	PIQSGNNKH	F	---	SGSSN	---	---	---	---	---	---	ISQN 617
Db	1573	---	TGGKKPR	PEASGSSR	DSLISQS	1595						

RESULT 7

Q7RG05 PRELIMINARY; PRT: 1268 AA.

ID Q7RG05

AC Q7RG05;

DT 01-MAR-2004 (TtEMBLrel. 26, Created)

DT 01-MAR-2004 (TtEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TtEMBLrel. 26, Last annotation update)

DE Eukaryotic translation initiation factor 3 subunit 10.

DE

GN Name=PY04546;

OS Plasmodium yoelii yoelii.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

NCBI_TaxID=73239;

UN [1]

RN

RP SEQUENCE FROM N.A.

RC STRAIN=17XNL;

RX PubMed=12369865; DOI=10.1038/nature01099;

RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,

RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,

RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,

RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feidblyum T.V.,

RA Cho J.K., Quackenbush J., Sedegham M., Shoabi A., Cummings L.M.,

RA Florens L., Yates F.R. Iii, Raine J.D., Sinden R.E., Harris M.A.,

RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,

RA

RESULT 8

Q9SI41 PRELIMINARY; PRT; 1738 AA.

AC Q9SI41;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative chromodomain-helicase-DNA-binding protein.

GN Name=At2g13370;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

SEQUENCE FROM N.A.

RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,

RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,

RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,

RA Fraser C.M., Venter J.C.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

SEQUENCE FROM N.A.

RA Town C.D., Kaul S.;

RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

RL -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -1- SIMILARITY: Contains 2 chromo domains.

DR EMBL; AC007209; AAD28668.1; -.

DR PIR; C84507; C84507.

DR GO; GO:0000785; C:Chromatin; IEA.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0005624; F:ATP binding; IEA.

DR GO; GO:0003682; F:chromatin binding; IEA.

DR GO; GO:0004386; F:helicase activity; IEA.

DR GO; GO:0006333; P:chromatin assembly or disassembly; IEA.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR011545; DEAD/DEAH N.

DR InterPro; IPR001650; Helicase C.

DR InterPro; IPR000330; SNF2_N.

DR Pfam; PF00385; Chromo; 1.

DR Pfam; PF00271; Helicase C; 1.

DR Pfam; PF00176; SNF2 N; 1.

DR SMART; SM00298; CHROMO; 1.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICC; 1.

DR PROSITE; PS50013; CHROMO_2; 2.

KW DNA-binding; Helicase; Nuclear protein.

SQ SEQUENCE 1738 AA; 198695 MW; CES0069DDB8C3B6CD CRC64;

Query Match 5.08; Score 160; DB 2; Length 1738;

Best Local Similarity 17.38; Pred. No. 2.9; Mismatches 28; Gaps 28;

Matches 132; Conservative 104; Indels 286; Gaps 28;

QY 2 ADAPILFGSSLNFSHVQKHVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNM-----56

DB 1041 ADTVIFDSWNPQNDLQAMRAHRIGQEVNIVFVTSKSVSEETILERAQRWLDHL 1100

QY 57 -----RQKAVENLRSLTHALMGCAVLFKLDHFSHSETPDSCVSEQS 103

DB 1101 VIQKNAEGRLEKRETKGNSFNDKNSLALRFGAEELFK-----1141

QY 104 IMDGVIFEFSSILSKGEBNEVKLCIL-----LEAKHAQCTYSDDSTLFGEDH 152

DB 1142 -----DKNDESKKRLSLMNDDEILERAQVEEKI---TDTEHELLGAFK 1184

QY 153 I-KLSDSESPNI FMSKLLGGKPNMKYPSPTQPNRRKRVQYFEGSEASPKTGQGNNAKR 211

DB 1185 VANFCNAEDDGSFWSRWIKPDSVVTAEEALAPRAARNTKSYVDPSPD-----RTSKRK 1238

QY 212 KKASDDVTRVTPPVDVDDERKASGKHMGALSPKVILOQSSCKSSGTDGLDGNDAF 271

DB 1239 KKGSE-----PPEHTERSQKRRTKTEYFVPSFTLLBGTSAQVRGWSYGNLFRKDAQ 1288

QY 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLQV 318

Db 1289 RFYRTVMKGNHMQACIAEEVGGVVE-----AAPBAAQVELFDA-----1328

QY 319 LHLSDACTSNV--GNP-----LEYYIENHRIYEPATTFQ 351

Db 1329 --LIDGCKESVETGNFEPKGPVLDFGVPVKANELLRKRVQGLQLSKRISRYNDPISQFR 1386

QY 352 AFQIALSWIA-----ALLVKQILSH--KESLVANSELAF--KCSRVEV 391

Db 1387 ----VUSYLUKPSNWSGCCWNQIDDLARLLGLLYHFGWKEKRLDESGLTKIAPVEL 1442

QY 392 DYIYSILSCWKSIFLEHTQGLQDFCF-----GTNSKQSVVSTKLVNESLSG---ATVRDE 443

Db 1443 QHHETFLPRAPNLKERATALLEMLAAAGGONTNAKASRKNKSKVDNLNQFKAPADR 1502

QY 444 K-----INTK-----SMRNSSE-----DEE 458

Db 1503 RGKSGPANVSLSTKDGPRKTKAEPKLVKEEGEMSDGGEYEQFKQKMWECEDVLADE 1562

QY 459 CMTKEKCSHYSTATRDI--EKTISGIKKKYK---KOVOKLVQEHKEKQMBELLNMYADKKQ 513

Db 1563 IKTLGRQLRQTTSADLPKEKLVFKIRRYLEILGRRIIDAIVLEHEE-----DLYQDRM 1616

QY 514 KLETSKSVEAAVIRITCSRTSTQVGLKLLDHNYERKFDIEKSEKNECLSKLQMEHAV 573

Db 1617 TMRL-----WNYVSTFNSLSDGR-----LNQIYSKLK 1643

QY 574 KKLAEDEACWINRIKSWAAKLVCVPIQSGNNKHFSGSSNISQ 616

Db 1644 QKXEEEGV-----GPSHLNGSRNFOR 1665

RESULT 9

Q7SHZ4 PRELIMINARY; PRT; 4007 AA.

AC Q7SHZ4;

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein.

GN Name=NCU00658.1;

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariales; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OR74A;

RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,

RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,

RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,

RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M., Schulte U.,

RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Greenberg D.,

RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,

RA Roy A., Foley J., Naylor J., Thomann N., Barrett R., Gnerre S.,

RA Kamal M., Kamayssellis M., Mauceli E., Bielek C., Rudd S., Frigman D.,

RA Krysotova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,

RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,

RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,

RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,

RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,

RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;

RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."

RL Nature 010-0(2003).

CC -1- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.

DR EMBL; AABX01000001; EAA36562.1; -.

DR HSSP; P01096; 1HF9.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.

DR InterPro; IPR003900; KID_repeat.

DR InterPro: IPR001638; SBP_bac_3.
DR Pfam: PF02524; KID; 5.
DR PROSITE; PS01039; SBP_BACTERIAL_3; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 4007 AA; 453246 MW; 67CDF6EF1463612 CRC64;

Query Match 4.9%; Score 158; DB 2; Length 4007;
Best Local Similarity 20.1%; Pred. No. 11;
Matches 129; Conservative 108; Mismatches 254; Indels 150; Gaps 28;

QY 17 DVKHEVKEIKIESCSERTKIFRLYSVCTVEEKALILARQNMKNRAVENLRSILTHALLMW 76
DB 531 DIEHWRQVDTLKKEDSLSL-----QEBA---TRKTEGADQMSLARTLN----- 575

QY 77 GASVFLDKLDFHSHSTPDGVSFEQIMDGVIFHEFSSILSSKGGENEVKLCILLLEAK- 135
DB 576 -----DIMDTAAALGVPIKGADPHDIGATILEVFDNINLNY--EDTEQRLVQTEBAKD 627

QY 136 HAQGTYSDDTLFGEDIH-KLSDEES-----PNIFWSKLGGKNPMWKPYSDT 182
DB 628 KAEAAKESATRSRSRTSIQKLTSENKLNKEQDRQKAKADAFYQKL---KKKGWLEPTED 684

QY 183 P-----QNRKRQVQVFEGSEASPKTGDGNAKR-KKASDDVDTPRV-----TDPVDD 230
DB 685 PSASLALDQSLKQDQDRWESEKVALNQKLTNLNKSIEBANVALKDKMLQVERDSLDL 744

QY 231 DERKASGDHMGALLES--PKVITLQSSCKSGTDTLGDNDAGFLYSMGSHIGIPEDML 288
DB 745 QORQE-----GDIKSLNQVQLDLKQKASKA-----SAGADLKDIOQLRL 784

QY 289 ASQDWGKIPDESQRRLHTVLKPKMAKLCQVLHLSLDACTSMVGNFLEVYVNIENRIYEPAT 348
DB 785 ENK-----SLEDQQR-----VQAEKEVLQOQLSOTKARLEKVTETLKNASQKMDLETQRN 836

QY 349 TFOAFQIALSWIAALLVKOL-LSHKESL-----VRANSELAPKCSREVEV----- 391
DB 837 EWSKAKVKLEESIRLKEVDLARKNSAAPTIVEIKEPNIASITVQELMKEQKQLP 896

QY 392 ----DYIYSILCMKSL-----FLEHTQGLQDFCGTNSKQSVVSTKLIVNESLSGATVRDE 443
DB 897 LTARDFVQGLVAKLKDQAQGHKDLDAEDVIRECIG-----LLNRVLK-EPVFE 944

QY 444 KIN-----TKSWNSSEDE--CMTEKRCSHYSTATRDIEKTIISGKKYKQVQ 491
DB 945 KDKEMLTQNLNEALRKGEDEQSAQLQSIIDRYEKESEKEV-----KVHDEIK 1000

QY 492 KLQVEHEEKQWELLN-----MYADKKOKLETSKSVAAVIRITCSRTSTOV---G 538
DB 1001 KLQVEITEKTSANKQOERDMLKESYEQIKNLNADHKKAAAEKVKHQNELTQLRKDG 1060

QY 539 DLKLLDHNRYERKFEIKSEKNECLKSLEQMHVAKKLAED 579
DB 1061 DLK--ETNLLQKLDTLAQONESERNRLQADYAAEKAKLTKD 1099

RESULT 10
Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein pf08_0127.
GN Name=pf08_0127;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51332.1; -.

KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Query Match 4.9%; Score 157; DB 2; Length 1245;
Best Local Similarity 18.1%; Pred. No. 2.8;
Matches 125; Conservative 111; Mismatches 259; Indels 196; Gaps 27;

QY 12 LNFSDHVKHVEKIKIESCSERTKIFRLY-----SVCTVEEKALILARQNMKNRAVENLN 66
DB 1 MNYDEDIKRRKKKTEKLTNTKESLSLYSNNKILSIKNNKLLKOKNNKIKQMKN-- 58

QY 67 RSLTHALLMWGASVLPDKLD-----HPHSSSTPDGVSFEQIMD----- 106
DB 59 -----KYDQDQGTIKHKGKNIISDDKDFDEITNLNKKKKKKCKMDTMVK 104

QY 107 -----GVIHEFSSILSSKGGENEVKLCILLLEAKHAQGTYSDDSTLF 148
DB 105 KKKKMKQNQYKDKNYGVVMDDEIID--NEERIENNVNNDVHPYPTNDSK--- 159

QY 149 GEDHIKLSDESPNIFWSKLGGKNPMWKPYSPTPQNRKRQVQVFEGSEASPKTGDGNA 208
DB 160 -----SDEEYSSVFMFKLKN-----DNEREVKKKKKKKSGTGNHKNDDHND 206

QY 209 KKRKASDDVDTPRVDDDERKASGDHMGALLES PKVITLQSSCKSGTDTLGDGN 268
DB 207 NDDNDDDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDN 247

QY 269 DAFGLYSMGSHIGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLC-----QVLHLSDA 324
DB 248 DDDNIKKLVH-SGDQDDHHVDDSSSVQDSYQ---YYALKDDMYKLSNENKEKIEPADL 303

QY 325 CTSWGNF-----LEVYVNIENR-----IYEEPATTFQAFQIALSWIAALLLVK 366
DB 304 MYSN-NNFQOGLQKLDLF-LENEQNKKGKGLSLFEE-----QKENTTLGYVQNVKLL 356

QY 367 QILSHKESLVRANSELAF-----KCSRVEVDYIYSILCMKSLFLEHTQGLQDFCGTNS 421
DB 357 NKWKAYNVITQNSKVEFGQKRGQGEDFVNKRL---VKRPLKKEKIKIPEMYDGHE 413

QY 422 KQSVV---STKLIVNESLSGATVRDEKINTKSMNSSEDEBCEMTEKRCSHYSTATRDIEKT 478
DB 414 ESGVIDDDNDDNINNI---LKRHKVDQTHLNGYGEQNEVYNNK-----EKK 457

QY 479 ISGIKKKYKQVQKLVQEHHEEK-KMELLNMYADKKOKLETSKSVAAVIRITCSRTSTOV 537
DB 458 I--IKREHKEEDFTIYNFEEIKMNL-----KSKMLHSED----- 493

QY 538 GDLKLLDHNRYERKFEIKS-EKNECLKSLEQMHVAKKLAEDAEACHINRIKSWAAKLKV 596
DB 494 ----LKNKEKRRDELKTAQLKLLSIENRRKNYKKTIKSKSYRKYLRKEKEKEBEKI 549

QY 597 CVPIQSGN-----NKHFGSSSNISQN 617
DB 550 MKKLYSEHPDLVKDLNMYEYAEKKNLNN 580

RESULT 11
Q9JK25 PRELIMINARY; PRT; 1320 AA.
AC Q9JK25;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CLIP-170.
GN Namescyl1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=21185938; PubMed=11290329; DOI=10.1016/S0092-8674(01)00288-4;


```

Db 194 -----LEDSDVIENSRKIA-----KVTNANESEVSHHKSDDHLLQIDQ 233
Qy 264 TLGNDAGFLYSGSHISGIPEDMLASQDW-GKIPDESQRRLHTVLKP-----KMAKLCQV 318
Db 234 VFNKEDLVGRFSSGASPAKLHEPTVIDEDSGDGIYNNNEVNTSHMTMTSKMRGNESQLSRI 293
Qy 319 LHLSDAC-----TSMVGNFLEYVLENHRIYEESPATTTQAFQIALSWTAALLVKQILS 370
Db 294 IKITDEVVDVYTPORYTEBPVAKSP-INOQHRIKENVINNTESKE-----TISPLKN 343
Qy 371 HKESLVRANSFELAFKCSRVE-----VDYIYISILSCMKSLF-LHETQGLQF 414
Db 344 HMLSVDNDQENHVSNNMEDENSDGFNELEDEPTINFLLSPNS--KPVFSLDHIKKVQD 401
Qy 415 DCFG-----TNSKQSVVSTKLVNESLSGAT-----VRDEKIN----- 446
Db 402 DHFKEVANLEEVNKNQELK-----FSEELSATNNKELIYDQKIKELKQDKKKLIANNENL 458
Qy 447 -----TKSMR-----NSSDEDECMTEKRCSHYSTATRDIEKTI----- 479
Db 459 LLIQLKGNRELASMTKALRIKENTVTQLESRLSKSKSYESTANELESVIEEGNTUREQ 518
Qy 480 -----SGIKKKYKQVQKLVQHEBEKKME-----LLNMVADKKQKLETSKSVAAV 525
Db 519 IKRLEKSVEDKSIQSEYEQIKLNDIEIKKOTEISLTDANIDYNIKVENLLSEKEEL 578
Qy 526 IRITCSRTSTQVGLKLLDHNHYERKFDIEKSEKNECLKSLEQMHVAKKLADEACWIN 585
Db 579 L-----TETGRLGRENGL-----EINSKQELLEELDLENLAKDKIMTILE-----N 622
Qy 586 RIKSWAAKLKVCVPIQSGNNKHFSGSSNISQ 616
Db 623 TLDSTKTOEIK--QVMTEKNEKLLTKIDNLTE 650

```

RESULT 14

```

Q7RRR9 PRELIMINARY; PRT; 2757 AA.
AC Q7RRR9
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Rhopty protein.
GN Name=Py00649;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNLI;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabai A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser F.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000178; EAA17609.1; -.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
DR PROSITE; PS00636; DNAJ_1; UNKNOWN_1.

```

```

SQ SEQUENCE 2757 AA; 324332 MW; 55AD73DE2B9EFA37 CRC64;
Query Match 4.8%; Score 154; DB 2; Length 2757;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 87; Conservative 69; Mismatches 159; Indels 124; Gaps 15;
Qy 245 ESPKVTITLOSSCKSSGDTGLDGNDAFLYSGSHISGIPEDMLASQDWGKIPDES--QR 302
Db 569 ENKVKILETKIQK-----DLFNQYL-----KIIDENIYN 598
Qy 303 RLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVLENHRIYEESPATTTQAFQIALSWIAA 362
Db 599 KLKLELKEKIKNISDKNEYIKKAVD-----LKKIENNANNAYIDELAKTSPYQVP----- 647
Qy 363 LLVKQILSKHESLVRANSELAPKCSR-----EVDYIYISILSCMKSLFLEHTQGLQFDCF 419
Db 648 -----EHLKSTDTIYSTIKSELQIYEDDIDKLYNELSSI-----VQENDINVED 693
Qy 420 NSKQSVVSTKLNVN-----ESLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRD 474
Db 694 KTKLDDLQSKIDNVYSKIQNMTATVESHILTNIETNKNKLSL--TIVEIKKIYGEISKD 751
Qy 475 IEKTIISGIKKKYKQVQKL-----VQHEBEKKWELLNMYADK-----KQK 514
Db 752 LNKTLDFKNKEKELSNKINDYAKENDQLNVYKSKISEIRNHYNDQISIDNTKEEBAQN 811
Qy 515 LETSKSVEAAVIRITCSRTSTQVGLKLLDHNHYERKFDIEKSEKNECLKSLEQMH----- 569
Db 812 YDQSKY-TKTISEKDETSKTINEVKNMKDEFLSKVDKYINFENCKENVDSHETQFTE 870
Qy 570 -----EVAKKKLA-----BDEACWINRIKSWAAKLKVC-----V 598
Db 871 LTNKIKAEVSDKLSIYENKFNDSKSLINETNNSIEKEYQNITLKKVDYIVKCESTKE 930
Qy 599 PIQSGNNKHFSGSSNISQ 617
Db 931 SIKNFHNKQITLKEKLNQN 949

```

RESULT 15

```

Q22342 PRELIMINARY; PRT; 1330 AA.
AC Q22342
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein T08A9.1.
GN Name=T08A9.1; ORFNames=T08A9.1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Pauley A.;
RT "The sequence of C. elegans cosmid T08A9."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 13:13:50 ; Search time 63 Seconds
(without alignments)
3812.352 Million cell updates/sec

Title: SEQ3_D1219_COPY_650_1270
Perfect score: 3217
Sequence: 1 RADAFILGSLNPSHDVXH.....SGNNKHFSSNISQNPDPV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3213	99.9	2001	4	AAB20062 Arabidops
2	3211	99.8	2001	4	AAB20063 Arabidops
3	163.5	5.1	1465	8	ADN22447 Bacterial
4	160.5	5.0	842	8	ADG66710 Novel hum
5	160.5	5.0	1739	6	AAE36105 Human chr
6	160.5	5.0	1739	6	AAE36106 Human chr
7	160.5	5.0	1739	8	ADK60208 Angiogene
8	160.5	5.0	1739	8	ADL12347 Human ste
9	160.5	5.0	1739	8	ADK60509 Angiogene
10	160.5	5.0	1739	8	ADP73132 Angiogene
11	157	4.9	1015	8	ADN20557 Bacterial
12	154	4.8	954	4	AAU14615 Novel bon
13	153.5	4.8	2228	7	ABR61599 Human gol
14	153.5	4.8	2230	6	ABU07445 Protein d
15	153.5	4.8	2230	7	ABR61600 Human gol
16	153.5	4.8	2250	7	ABR61601 Human gol
17	153.5	4.8	2252	7	ABR61602 Human gol
18	149	4.6	1881	5	ABP73809 Candida a
19	147.5	4.6	823	4	ABU53074 Intracell
20	147.5	4.6	875	4	ABU53078 Intracell
21	146	4.5	1455	7	ADG80783 Microsate
22	146	4.5	1948	4	ABG21233 Novel hum
23	146	4.5	1959	5	ABU10604 Human nov
24	145	4.5	962	3	AAU18211 Plasmodiu
25	144.5	4.5	3210	6	ABU07438 Protein d

ALIGNMENTS

RESULT 1

AAB20062
ID AAB20062 standard; protein; 2001 AA.

AC AAB20062;

XX 04-NOV-2004 (revised)

DT 23-APR-2001 (first entry)

XX Arabidopsis thaliana silencing gene-encoded protein.

XX Gene silencing; silencing gene; MOM.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT Region 177..350

FT Peptide /note= "internal repeat region"

FT Binding-site 362..367

FT /note= "nuclear localization motif"

FT /note= "ATP/GTP-binding motif"

FT /note= "region of homology to ATPase/helicase family

FT /note= "SWI2/SNF proteins"

FT Misc-difference 705

FT Peptide /note= "Lys in ecotype Columbia"

FT Peptide 832..838

FT /note= "nuclear localization motif"

FT Peptide 858..862

FT /note= "nuclear localization motif"

FT Domain 995..1015

FT /note= "predicted membrane-spanning domain"

FT Misc-difference 1219

FT /note= "Asp in ecotype Columbia"

FT Region 1462..1672

FT /note= "internal repeat region"

FT Region 1848..1894

FT /note= "internal repeat region"

FT Domain 1899..1941

FT /note= "actin-binding domain"

XX WO200100801-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-BP005761.

XX

PR 23-JUN-1999; 99GB-00014623.
XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX WPI; 2001-137952/14.
DR N-PSDB; AAA89353, AAA89354.
XX
PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 1; Page 32-39; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana protein MOM, which
CC controls gene silencing, particularly gene silencing. It is encoded by a
CC novel silencer gene (see AAA89353) of A. thaliana ecotype Zurich. The
CC protein exhibits partial similarity with ATPase/helicase proteins of the
CC SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
CC compared with the protein (see AAB20063) encoded by the silencer gene of
CC wild-type A. thaliana ecotype Columbia. Gene silencing is useful as a
CC molecular tool for regulating gene expression
CC
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
SQ Sequence 2001 AA;

Query Match 99.9%; Score 3213; DB 4; Length 2001;
Best Local Similarity 99.8%; Pred. No. 3e-267;
Matches 620; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADAFILGSSLNPNSHDVHVKIEKIESCSERTKIFRLYSVCTVEKALILARQNRQK 60
Db RADAFILGSSLNPNSHDVHVKIEKIESCSERTKIFRLYSVCTVEKALILARQNRQK 709

QY 61 AVENLNRSILTHALLMWGASYPDKLDHFSSETPGSGVSFEQSDMGVIEHFSILSSKG 120
Db AVENLNRSILTHALLMWGASYPDKLDHFSSETPGSGVSFEQSDMGVIEHFSILSSKG 769

QY 121 GEENEVKCLLLEAKHAQCTYSSTDLFGEDHILKLSDESPNIFWSKLGGNPNWKYPS 180
Db GEENEVKCLLLEAKHAQCTYSSTDLFGEDHILKLSDESPNIFWSKLGGNPNWKYPS 829

QY 181 DTPQNRKRVQVFEQSEASPKTGDGNNAKRRKASDDVTPDPRVTPPPVDDDERKASGKH 240
Db DTPQNRKRVQVFEQSEASPKTGDGNNAKRRKASDDVTPDPRVTPPPVDDDERKASGKH 889

QY 241 MGALSPKPVITLQSSCKSGTGTLDGNDAPGLYSNGSHISGIPEDMLASQDWGKIPDS 300
Db MGALSPKPVITLQSSCKSGTGTLDGNDAPGLYSNGSHISGIPEDMLASQDWGKIPDS 949

QY 301 QBRLLHTVLKPKWAKLCQVHLSDACTSMVGNFLEVYENHRIYERPAITFOAFQALSWI 360
Db QBRLLHTVLKPKWAKLCQVHLSDACTSMVGNFLEVYENHRIYERPAITFOAFQALSWI 1009

QY 361 AALLVKQILSHKESLVRANSELAFCRSRVEVDYIYSLSCMKSLFLEHTQGLQFCFGTN 420
Db AALLVKQILSHKESLVRANSELAFCRSRVEVDYIYSLSCMKSLFLEHTQGLQFCFGTN 1069

QY 421 SKQSVVSTKLVNESLSGATVREKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKITS 480
Db SKQSVVSTKLVNESLSGATVREKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKITS 1129

QY 481 GIKKKYKQVQKLVQHEHEKKMELNNYADKKQLETSKVEAAVIRITCSSTSTQVGLD 540
Db GIKKKYKQVQKLVQHEHEKKMELNNYADKKQLETSKVEAAVIRITCSSTSTQVGLD 1189

QY 541 KLLDHNRYERKFDIEIKSEKNECLKSLQEQMDHVAKKLADEACWINKRIKSWAAKLKVCVPI 600
Db KLLDHNRYERKFDIEIKSEKNECLKSLQEQMDHVAKKLADEACWINKRIKSWAAKLKVCVPI 1249

QY 601 QSGNKNHFGSSNISQNPADV 621

Db 1250 QSGNKNHFGSSNISQNPADV 1270

RESULT 2
AAB20063
ID AAB20063 standard; protein; 2001 AA.
XX
AC AAB20063;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene-encoded protein.
XX
KW Gene silencing; silencing gene; MOM.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Region 177..350
FT /note= "internal repeat region"
FT Peptide 362..367
FT /note= "nuclear localization motif"
FT Binding-site 460..467
FT /note= "ATP/GTP-binding motif"
FT Region 479..719
FT /note= "region of homology to ATPase/helicase family
FT SWI2/SNF proteins"
FT Misc-difference 705
FT /note= "Met in ecotype Zurich"
FT Peptide 832..838
FT /note= "nuclear localization motif"
FT Peptide 858..862
FT /note= "nuclear localization motif"
FT Domain 995..1015
FT /note= "predicted membrane-spanning domain"
FT Misc-difference 1219
FT /note= "Glu in ecotype Zurich"
FT Region 1462..1672
FT /note= "internal repeat region"
FT Region 1848..1894
FT /note= "internal repeat region"
FT Domain 1899..1941
FT /note= "actin-binding domain"

XX WO200100801-A2.
XX
XX 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-EP005761.
XX
XX 23-JUN-1999; 99GB-00014623.
XX
XX (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
DR N-PSDB; AAA89385.
XX
PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 8; Page; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana protein MOM, which
CC controls gene silencing, particularly gene silencing. It is encoded by a
CC novel silencer gene (see AAA89385) of A. thaliana ecotype Columbia. The
CC protein exhibits partial similarity with ATPase/helicase proteins of the
CC SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
CC compared with the protein (see AAB20062) encoded by the silencer gene of
CC Arabidopsis thaliana ecotype Zurich. Gene silencing is useful as a

Handwritten signature

CC molecular tool for regulating gene expression. Note: the present sequence
CC is not shown in the specification but is derived from the A. thaliana
CC mutant silencer gene protein sequence given in AAB20062
XX
SQ Sequence 2001 AA;

Query Match 99.8%; Score 3211; DB 4; Length 2001;
Best Local Similarity 99.8%; Pred. No. 4.5e-267;
Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RADAFILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 60
DB |||||
QY 61 AVENLNRSLSLTHALLMGCASYLFDKLDHFSSETPDGSGVSEQIMDGVIEHFSILSSKG 120
DB |||||
QY 121 GEENEVKLCILLLEAKHAQGYSSDSTLFGEDHKLSDSESPNIFWSKLLGGKRNPMKYP 180
DB |||||
QY 770 GEENEVKLCILLLEAKHAQGYSSDSTLFGEDHKLSDSESPNIFWSKLLGGKRNPMKYP 829
DB |||||
QY 181 DTPQRNKRVOYFEGSEASPTGDNNAKRRKASDDVTPDPPVDDDERKASGKH 240
DB |||||
QY 830 DTPQRNKRVOYFEGSEASPTGDNNAKRRKASDDVTPDPPVDDDERKASGKH 889
DB |||||
QY 241 MGALESFKVITLQSSCKSGTDLGNDAGFLYSMSGSHISGIPEDMLASQDWGKIPDES 300
DB |||||
QY 890 MGALESFKVITLQSSCKSGTDLGNDAGFLYSMSGSHISGIPEDMLASQDWGKIPDES 949
DB |||||
QY 301 QRLHVTLPKMAKLCQVLSLDACTSMVGNFLEYVIEHRIYEEPAITTFQAFQIALSWI 360
DB |||||
QY 950 QRLHVTLPKMAKLCQVLSLDACTSMVGNFLEYVIEHRIYEEPAITTFQAFQIALSWI 1009
DB |||||
QY 361 AALLVQIILSHKESLRANSELAFKCSRVEDVYIYLSCKMSLFLHTQGLQDFCGTN 420
DB |||||
QY 1010 AALLVQIILSHKESLRANSELAFKCSRVEDVYIYLSCKMSLFLHTQGLQDFCGTN 1069
DB |||||
QY 421 SKQSVVSTKLNVESLSGATVRDEKINTKSMRNSSEDEECMTERKCSHYSTATDIEKTIS 480
DB |||||
QY 1070 SKQSVVSTKLNVESLSGATVRDEKINTKSMRNSSEDEECMTERKCSHYSTATDIEKTIS 1129
DB |||||
QY 481 GIKKKYKQVQKLVQHEEKMBELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 540
DB |||||
QY 1130 GIKKKYKQVQKLVQHEEKMBELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 1189
DB |||||
QY 541 KLDDHNYERKFDRIKSEKNECLSLQMDHVDVAKKLAEDACWINRIKSWAAKLVKCVPI 600
DB |||||
QY 1190 KLDDHNYERKFDRIKSEKNECLSLQMDHVDVAKKLAEDACWINRIKSWAAKLVKCVPI 1249
DB |||||
QY 601 QSGNNKHFSGSSNISQNPV 621
DB |||||
QY 1250 QSGNNKHFSGSSNISQNPV 1270
DB |||||

RESULT 3
ID ADN22447
XX ADN22447 standard; protein; 1465 AA.

AC ADN22447;

XX 02-DEC-2004 (first entry)

XX Bacterial polypeptide #5100.

XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polypeptide.

OS Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOV/) CAO Y.
PA (HINKLE/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 5100; 122pp; English.

CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: the sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 1465 AA;

Query Match 5.1%; Score 163.5; DB 8; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.00044;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADATILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNQRNK 61
DB |||||

QY 799 ADTVIIFDSDWNPQNDLQMSRAHRIGQTKTNYIKLVTKGSVEE-----EI 845

QY 62 VENLNRSLSLTHALLMGCASYLFDKLDH-----FHSSETPDGSGVSEQIMDGV 108
DB |||||

QY 846 VERAKKLV-----LDHLVIQRMDTTGKTVLSKNATASGSPFDKQLSAI 891

QY 109 IHEFSSI--LSSKGGRENEV-----KCLLLAKHAQGYSSDSTLFGEDHKLSDSESPN 160

QY 892 L-KFGAVELFKKEGEQEPEVDIDRLMGAETREAEVEEMKENELLSFKYANFAIDEE 950

QY 161 PNIF-----WSKLGKGNPMKYPDTPQNRKV-----Q 191

QY 951 KOIAAATDEWAAII-----PEEDRNRIIEERMKELAEMLNAPQRKQPIQ 997

KW cytostatic.
XX Homo sapiens.
OS
PN WO200298899-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017466.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX
PA (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
PI
XX
XX WPI; 2003-156840/15.
XX
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.
XX
XX Claim 13; Page 185-192; 278pp; English.
XX
XX The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human chromodomain
CC helicase DNA binding protein
XX
SQ Sequence 1739 AA;

Query Match 5.0%; Score 160.5; DB 6; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.001;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY 2 ADAPILFGSSLNPSHDVKEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVVIFDSDWNPQNDLQARAHRIQKKQVNIYRLVTKGTVEEILIERAKKMWLDHL 938
QY 50 ILARQNMQRNKAVERN-----LNRSILTHALLMMGASYLFDKLDHFSHSTPPDSGVSF 100
Db 939 VIQRMDTTCRTILENNSGRSNPNFKBELTALFKCAEDLPKELE-----GEES 988
QY 101 EQSIMGVTHFESSILSSKGGENEVKLCLELAKHQAQTYSSDSLTFGEHDHKLSDRES 160
Db 989 EPQEMD-----IDEILRLAETRENEVSTATDILL-SQPKVANFATMEDEEEL-----BER 1038
QY 161 PNIWFSLKLGKPNMKYPSTDPQNRKRVQYFE-----GSEASPKTG 203
Db 1039 PHKWDDEII-----PEQRKKVEEERQKELEIYMLPRISSTYKAQTND 1085
QY 204 DGGNAKRRKASDDVTPDPRVTPDPPVDDDERKASGKDHMGALSPKVITLQSSCKSGTGD 263
Db 1086 DSDTESKRQAQRSSASESETD---SDDDKK-----PKRGRPRSVRKDLVEG 1130
QY 264 TLQGN-----DAPGLYSMGSHTIGIPEDML--ASQDWKIPDESQRLLHTVLKPRVAKLC 316
Db 1131 FTDAEIRRFITKAYKF-----GLPLERLECLARD-ABLVDKSVADL-----KRLG 1174

QY 317 QVHLSDACTSMVGNFLEVYENHRIYEBPATTFOAFQIALSWIAALLVQIILSHKESLV 376
Db 1175 ELTH--NSCVSAQMEYEEQLENASEGKPGKR-RGPTIKISGV-QVNVKSIITQHEEFE 1230
QY 377 RANSEL-----APKCSRVEVDYIYSILSCMSLPLEHTQGLQDFCFGTNSK 422
Db 1231 MLHKSIPVDPPEKKKYCLTCRVKAAHFDVWGVEDDSRLLLGIYEHGYG-NWSLIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMNSSE-----DRECHTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPOGKQLQTRADYLLKLLKGLKKGAVTGGEEAKLKR 1342
QY 465 -----CSHYSTATRDIENTISGIKKIKYKKQVQKLVQEHHEKK 501
Db 1343 KPRVKKENVPRLKBEHGIELSPRHSNDPSEBEVGDGDKLEKSPMKCKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KQKLETSKVEA-----AVIRITCSTSTQVGD--- 539
Db 1402 EKOMSSRKDKGDKERKKSKDKKPKSGDAKSSSKSRKSGQGPVHITAGSEPVPIGEDE 1461
QY 540 -----LKLLDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLDDQETPSICKERMPVKALKQLDKPKGLNVQEQLEHTRNCLLKI GDRIAECLKAYS 1521
QY 565 -----LEQMHDVAKKLADEACWINRIKSWAAKUKVCV 598
Db 1522 DQEHKLMRRNLWIFVSKTFEFDARKLHLYKMAHKRQSEEE-----EQKKKDDV 1572
QY 599 PTQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGGKDFRPEASGSRDSLISQS 1595

RESULT 6
ID AAE36106 standard; protein; 1739 AA.
AC AAE36106;
XX 26-JUN-2003 (first entry)
XX Human chromodomain helicase DNA binding protein (CHD) #4.
XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; Gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Domain 281..327
FT Domain /note= "Chromodomain"
FT Domain 375..433
FT Domain /note= "Chromodomain"
FT Domain 487..768
FT Domain /note= "SNEF2N domain"
FT Domain 831..905
FT Domain /note= "Conserved C-terminal domain"
XX WO200298899-A2.
XX 12-DEC-2002.
XX 03-JUN-2002; 2002WO-US017466.
XX 05-JUN-2001; 2001US-0296076P.
XX 10-OCT-2001; 2001US-0328605P.
XX 22-OCT-2001; 2001US-0338733P.
XX 15-FEB-2002; 2002US-0357253P.
XX 15-FEB-2002; 2002US-0357600P.
XX

PA	(EXEL-) EXELIXIS INC.	
XX	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;	
PI	Lioubin MN;	
XX		
XX	WPI; 2003-156840/15.	
XX	Identifying a candidate p53 pathway-modulating agent as therapeutic	
PT	targets for disorders related to defective p53 function e.g. cancer by	
PT	contacting an assay system having purified CHD polypeptide or nucleic	
PT	acid, with a test agent.	
XX		
PS	Claim 13; Page 192-199; 278pp; English.	
XX		
CC	The present invention relates to a method for identifying candidate p53	
CC	pathway modulating agents. The method involves contacting an assay system	
CC	comprising purified chromatin organisation modifier (chromo) domain	
CC	helicase DNA binding proteins (CHD), nucleic acids, their functionally	
CC	active fragments or derivatives, with a test agent under conditions	
CC	where, but for the presence of the test agent, the system provides a	
CC	reference activity. The methods are useful for identifying modulators of	
CC	the p53 pathway as therapeutic targets for disorders associated with	
CC	defective p53 function, such as angiogenic disorders, apoptotic disorders	
CC	or cell proliferative disorders, e.g. cancer. The modulators are useful	
CC	as research reagents, diagnostics and therapeutics. The invention is also	
CC	useful in gene therapy. The present sequence is human chromodomain	
CC	helicase DNA binding protein	
SQ	Sequence 1739 AA;	
Query Match	5.0%; Score 160.5; DB 6; Length 1739;	
Best Local Similarity	18.9%; Pred. No. 0.001;	
Matches	152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;	
QY	2 ADATILFGSLSPSHDVHKEVKEIKIESCSERTKIFLYSVCTVEKA-----L 49	
DB	879 ADTVVIFDSWPNQDLQAARAHRIQOKKQVNIYRLVTKGTVEEIIIRAKKQVLDHL 938	
QY	50 ILARQNMQRKQAVEN-----LNRSILTHALLMAGASYLFDKLDHPSHSETPPDSGVSF 100	
DB	939 VIQRMDTTGRTILENNSGRNSNPFNKEELTALKEGAEDLFKELE-----GEES 988	
QY	101 EQSIMGVTHFSSILSGGGENEVKLCILLEAKHAQGTYSDDSTLFGEDHDKLSDEES 160	
DB	989 EPQEMD-----IDEILRLAETRENEVTSATDELL-SQFKVANFATWEDEEEL-----EER 1038	
QY	161 PNIFFSKLGGKNPMWKYPSDTPQRNRKVQYFE-----GSEASPKTG 203	
DB	1039 PHKDWEIIL-----PEQRKKVEEERQKELEEIYMLPRISSYTKKATNDS 1085	
QY	204 DQGNAKRKKASDDVTPDPVDDDERKASGDHMGALBSPKVITILQSSCKSGTGDG 263	
DB	1086 DSDTESKQAOSSASESED---SDDDK-----PKRGRPRSVRKDLVEG 1130	
QY	264 TLQGN-----DAFLYSGMISHGIPEDML--ASQDWKIPDESQRRLHTVLKPKVAKLC 316	
DB	1131 FTDAEIRRETIKAYKF-----GLPLERLECLARD-AEIVDKSVADL-----KRLG 1174	
QY	317 QVILHLSDACTSMVGNFLEYVNIENHRIYEEPATTFQAFQIALSWIALLKQILSHKESLV 376	
DB	1175 ELIH--NSCVSAMQVEEQLKENASGKGPGRK-RGPTIKISGV-QVNVKSIQHEEPEF 1230	
QY	377 RANSEL-----AFKCSRVEVDYIYISLSCMKSLPLEHTQGLQDFCFNTSK 422	
DB	1231 MLHKSIPVDPPEKKYCLTRCVKAHFDEWGVDDSDRLLLGIYEHYG-VNWEILKTDFE 1289	
QY	423 QSVSTVKLVNESLSGATVDEKINTKSMNSP-----DEECWTEKR 464	
DB	1290 -----LKLTDKILPVT--DKPQGLQTRADYLLKRLKGLKFKGAVTGGBEAKLKR 1342	
QY	465 -----CSHYSTATDIETKISGKIKKKYKQVKLVQREHEKK 501	
DB	1343 KPRVKENKVPRLKEEHGIELSSPRHSDNPSEGEVKKDGLKSPMKKKQK-KXENKKNK 1401	

QY	502 MELLNNYADK-----KQKLETYSKVEA-----AVIRITCSRTSTQVGD----	539
DB	1402 EQQMSRRDKDKEGDKERKSKDKKKEPKSGDAKSSKSKRSQGPVHITAGSEPVPIGDEED	1461
QY	540 -----LKLDDH-----NYERKPD-----EIKSEKNECLKS--	564
DB	1462 DDLDOQETFSICKERMPVKALKQDKPGLNVQLEHTNCLLKIGDRIAECLKAYS	1521
QY	565 -----LEOMHDVAAKKLAEDACWINRIKSWAAKLKVCV	598
DB	1522 DOEHIKLMWRNLWIFVSKFTEFDARKLHLYQMAHKRSQEEE-----EQKKDDV	1572
QY	599 PIQSGNNKHF-----SGSSN-----ISQN 617	
DB	1573 ---TGKKPFRPEASGSSRDSLISQS 1595	
RESULT 7		
ID	ADK60208 standard; protein; 1739 AA.	
XX	ADK60208;	
DT	06-MAY-2004 (first entry)	
DE	Angiogenesis differentially expressed protein GS-P32.	
KW	vasotropic; antirheumatic; antiarthritic; hypotensive; antidiagonal;	
KW	antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;	
KW	angiogenesis; endothelial cell; diagnosis; tumor vascularization;	
KW	retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;	
KW	ovary hyperstimulation; psoriasis; endometriosis; restenosis;	
KW	angioplasty; cicatrization; peripheral vascular disease; hypertension;	
KW	vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;	
KW	ischemia; angina; myocardial infarction; chronic heart disease;	
KW	cardiac congestion; macular degeneration; osteoporosis.	
OS	Homo sapiens.	
XX		
PN	FR2836687-Al.	
XX		
PD	05-SEP-2003.	
XX		
PF	11-APR-2002; 2002FR-00004546.	
XX		
PR	04-MAR-2002; 2002FR-00002717.	
XX		
PA	(GENE-) GENE SIGNAL.	
PA	(ALMA/) AL MAHMOOD S.	
XX		
PI	Colin S, Schneider C, Al Mahmood S;	
XX		
DR	WPI; 2004-013912/02.	
XX	N-PSDB; ADK60458.	
PT	Compositions for diagnosing, prognosing and treating angiogenic disorders	
PT	including tumor vascularization and heart disease, comprise nucleic acid	
XX	or polypeptide differentially expressed in angiogenesis.	
PS	Claim 7; SEQ ID NO 84; 424pp; French.	
XX		
CC	The invention relates to a novel pharmaceutical composition active on	
CC	angiogenesis comprising an endothelial cell nucleic acid whose expression	
CC	is induced by an angiogenic factor and inhibited by an angiostatic agent	
CC	or its complement or fragment, a polypeptide sequence encoded by the	
CC	nucleic acid or its fragment, a molecule capable of inhibiting expression	
CC	of the nucleic acid or a molecule which binds to the polypeptide	
CC	sequence. The invention is used to diagnose, prognose or treat an	
CC	angiogenic disorder in a mammal, particularly a human. The disorder is	
CC	particularly tumor vascularization, a retinopathy, rheumatoid arthritis,	
CC	Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,	
CC	endometriosis associated with neovascularization, restenosis due to	

CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
CC aneurism, arterial stenosis, thrombophlebitis, ischemia, angina,
CC myocardial infarction, chronic heart disease, cardiac congestion or
CC macular degeneration due to age or osteoporosis. This sequence
CC corresponds to a protein encoded by a differentially expressed DNA used
CC in the composition of the invention.

XX
SQ Sequence 1739 AA;

Query Match 5.0%; Score 160.5; DB 8; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.001;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY 2 ADAPILFGSLNPSHDVHKVKEIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTWIFDSQNDLQAARAHRIQKQKQVNIYRLVTGTEETIERAKKQVLDHL 938
QY 50 ILARQNRQNKAVEN-----LNRLTHALLMGASVLFKLDHFSSETPDGVSF 100
Db 939 VIQMDTGTGRTILENNSGRNSPNFKBELTALFKGAEDLFKELE-----GEES 988
QY 101 EQSIMDGVIEHPSILSKGGENEVKLCLEAKHAQGYSSDSTLFGEDHKLSDSES 160
Db 989 EPOEMD-----IDEILRLAETRENEVSTATDELL-SQKVFANFATMEDEEL-----EER 1038
QY 161 PNIFWSKLLGGKPNMKYPSDTPQNRKRQVYFE-----GSEASPKTG 203.
Db 1039 PHKDWDEII-----PEQRKVEEERQKELEIYMLPRISSTKKAQTNDS 1085
QY 204 DGNNAKRRKASDDVTPDPRVDDDERKAGKHMGALSPKVIITLOSSCKSGTGD 263
Db 1086 DSDTESKQARQASASETED---SDDDKK-----PKRGRPRSRVKDLVEG 1130
QY 264 TLQGN-----DAFLYSMGSHISGPEMDL--ASQDWGKIPDESQRRLHTVLPKMAKLC 316
Db 1131 FTDAEIRRFYKAKKF-----GLPLERLECLARD-AEILVDSKADL-----KRLG 1174
QY 317 QVLHLSDACTSMVGNFLEYIENHRIYEEPAITFOAFIALSIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQVEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIQHEEPE 1230
QY 377 RANSEL-----AFKCSRVVDYIYSILSCMSKSLFLEHTQGLQFCFGTNSK 422
Db 1231 MLHKSIPVDPPEKKKYCLTCRVKAAHFDVWGVEDDSRLLLGIYEHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVREKINTKSMRNSSE-----DEECWTEKR 464
Db 1290 -----LKLTDKILPVET--DKPQGLQTRADYLLKLLKLGKLEKKGAVTGGBEAKLKR 1342
QY 465 -----CSHYSTATRDIETKISGKKYKQVQKLVQEHHEKK 501
Db 1343 KPRVKENKVPRLKEEHGIELSPRSDNPSSEGEVKKDGLKSPMKKQK-KKENKEN 1401
QY 502 MELLNMYADK-----KQLETSKSVBA-----AVIRITCSTSTQVGD--- 539
Db 1402 EKQMSRRKKEGDKERKSKDKKPKSGDAKSSKSKRSQGPVHTAGSEPVPIGEDD 1461
QY 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DLDLQETFTSTCKERMPPVKALKQLDKPGLNVQEOLEHTRNCLLKIGRIAECIKAYS 1521
QY 565 -----LEOMHDVAKKLADEACWINRIKSWAALKVCV 598
Db 1522 DQEHKLRNRLNIFVSKFTFEDARKLUYKWAHKRSQEE-----BQKKDDV 1572
QY 599 PISGNNKHFF-----SGSSN---ISQN 617
Db 1573 ---TGGKKPFRPEASGSSRDSLISQS 1595

RESULT 8
ADL12347

ID ADL12347 standard; protein; 1739 AA.
XX
AC ADL12347;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human steroid-induced C3A liver cell protein #14.
XX
XX Hepatotropic; Gene therapy; Wilson disease; liver disorder;
KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
OS Homo sapiens.
XX
XX US6673549-B1.
PD 06-JAN-2004.
XX
XX 12-OCT-2001; 2001US-00976594.
XX
XX 12-OCT-2000; 2000US-0240409P.
PR
XX
XX (INCY-) INCYTE CORP.
XX
XX Furness LM, Buchbinder JL;
XX
XX WPI; 2004-068610/07.
DR
XX
XX Combination useful for preparing a composition for treating liver
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
PT comprises cDNAs that are differentially expressed in response to steroid
PT treatment.

Disclosure; SEQ ID NO 76; 141pp; English.

CC The invention relates to a combination comprising cDNAs that are
CC differentially expressed in response to steroid treatment. Also included
CC are the following: a high throughput method for using a cDNA to detect
CC differential expression of nucleic acids in a sample; and a high
CC throughput method of screening molecules or compounds to identify a
CC ligand that specifically binds a cDNA. The sample is from a subject with
CC Wilson disease and comparison of a standard defines a stage of that
CC disease. The high throughput method of screening molecules or compounds
CC to identify a ligand that specifically binds a cDNA comprises: combining
CC the combination with molecules or compounds under conditions to allow
CC specific binding; and detecting specific binding between each cDNA and at
CC least one molecule or compound. The molecules or compounds are regulatory
CC proteins. The combination is useful for preparing a composition for
CC treating liver disorders associated with steroid therapy, e.g., cirrhosis
CC or hepatitis. The present sequence represents a human protein which is
CC differentially expressed in steroid-induced C3A liver cells. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 1739 AA;

Query Match 5.0%; Score 160.5; DB 8; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.001;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY 2 ADAPILFGSLNPSHDVHKVKEIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTWIFDSQNDLQAARAHRIQKQKQVNIYRLVTGTEETIERAKKQVLDHL 938
QY 50 ILARQNRQNKAVEN-----LNRLTHALLMGASVLFKLDHFSSETPDGVSF 100
Db 939 VIQMDTGTGRTILENNSGRNSPNFKBELTALFKGAEDLFKELE-----GEES 988
QY 101 EQSIMDGVIEHPSILSKGGENEVKLCLEAKHAQGYSSDSTLFGEDHKLSDSES 160
Db 989 EPOEMD-----IDEILRLAETRENEVSTATDELL-SQKVFANFATMEDEEL-----EER 1038
QY 161 PNIFWSKLLGGKPNMKYPSDTPQNRKRQVYFE-----GSEASPKTG 203

RESULT 12
AAU14615
ID AAU14615 standard; protein; 954 AA.
XX AC
XX AAU14615;
DT 24-OCT-2001 (first entry)
XX DE
XX Novel bone marrow polypeptide #14.
XX KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX OS
XX Homo sapiens.
XX WO200157187-A2.
XX PN
XX 09-AUG-2001.
XX PD
XX 05-FEB-2001; 2001WO-US0003782.
XX PF
XX 03-FEB-2000; 2000US-00496914.
XX PR
XX 20-JUN-2000; 2000US-00598075.
XX PR
XX 19-JUL-2000; 2000US-00620325.
XX PR
XX 30-NOV-2000; 2000US-0250683P.
XX XX
XX (HYSE-) HYSEQ INC.
XX PA
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
PI
XX WPI; 2001-488875/53.
XX DR
XX N-PSDB; AAS22920.
XX XX
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
PT Gene therapy.
PT
XX
PS Claim 10; Page 215-217; 392pp; English.
XX PS
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used in
CC the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein. The
CC nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins and
CC in assays to identify modulators of their expression and activity. The
CC anti-bone marrow protein antibodies and antagonists may also be used to
CC down regulate expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be
CC used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID)
XX
XX Sequence 954 AA;
XX SQ
Query Match
Best Local Similarity 4.8%; Score 154; DB 4; Length 954;
Matches 131; Conservative 105; Mismatches 233; Indels 224; Gaps 28;
QY 93 TPDGVSFQSGIMDGVTHFS-----SILSSKGGNEEVKLCLEAKHAQGYSSDSTLIF 148

Db 19 SPGAGSSLSAVGKG--RQSGARGPGVPEKEENKNEPKICRMVMVAPKADPT----- 68
QY 149 GEDHIKLSDEESPIFWSKLLGG-----KNPMKY-----PSTTPQ-----RNR 187
Db 69 GSEPAKASEKAPE---DTVDAGEGSRREAAKEPKKASALEEGSSDASQLEISEHM 125
QY 188 KRVQYFEGSEASPKTGDGNAKKRKKASDDVTDPRVTDP-----PVDDDERK 234
Db 126 KEPQLSDSIASDPKSPFHGLDFGRSRISEHLLDVLSPVLGGACRQAQOPLGIEDKDD 185
QY 235 ASGKHMGALLES-----PKVITLQSSCKS-----SGTDGTLT----- 266
Db 186 QSSQDELQSKQSGLEERYHRLSPPLPHEERAQSPRSLATEEPQGGPEGQEWKEAE 245
QY 267 -GNDAFGLYSM-----GSHISGIPEDMLA----- 289
Db 246 LGEDSAASLSLQSLQREQAQSPPAACEKKGQHSQAELGPGQEAEPPEEKVAVSPT 305
QY 290 -----SQDWGKIPDESQRRLHTVLKPKM-----AKLCQ----- 317
Db 306 PVSPEVRSTEPVAPPEQLSEAAKAMEEAQAQVLEQDQRHLLSKQEKMQQLREKLCQEE 365
QY 318 ---VLHL-----SDACTSMWGNFLEYVYENHRIYEEPAATTAQAFQIALSWIAALLVKQILSH 371
Db 366 EEEILRLHQKQESLSSLEERLQKAIE-----EEARMREESQRSLSWLRA-QVQSSTQA 419
QY 372 KESLVRANSELAPKCSRVEDVYIYILSCMK-----SLFLEHTQGL-----QDFCFQTNK 422
Db 420 DEDQIRAEQASLQKLEEE-----SQKAEASLEQKNQMLQLEKLEIEA-SEKSE 472
QY 423 QSVWST-----KLVNESLSGATVRDEKINTKSMRNSSEDECEMTEKRGSHYSTATPDI- 475
Db 473 QAALNAAKEKALQQLREQLGE--RKEAVATLEKEHSAB-----LERLCSLEAKHREV 525
QY 476 ---EKTISGKKYKQVQKLVQEHKEK-----KMLLNMYADKKQKLE----- 516
Db 526 SSLQKKIQEAQKQKBEAQKCGQVHRVHQS YHVAGYEHLSELSSLLREKQVEGEHER 585
QY 517 -----TSKSVAAAVIRITCSRTSTQVGDLLKLDHNVYERKPFDEIKSEKNEC 561
Db 586 RLDMKKEHQVMAKAREQYEAERKQRAELGLHTGELERLQRAHRELETVRQEHKR 645
QY 562 LKSLQMHQVMAKLADEACWINRIKSWAAKL 594
Db 646 LEDLARRHREKERKQLDLELDLETRAKDKVKARL 678
RESULT 13
ABR61599
ID ABR61599 standard; protein; 2228 AA.
XX AC ABR61599;
XX DT 15-JAN-2004 (first entry)
XX DE Human golgin-245 splice variant 1 polypeptide.
KW Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX OS Homo sapiens.
XX WO2003087403-A2.
XX PN 23-OCT-2003.
XX PD
XX PF 16-APR-2003; 2003WO-EP003958.
XX PR 16-APR-2002; 2002EP-00008553.
XX PR 16-APR-2002; 2002US-0372424P.
XX PA (EVOT-) EVOTEC NEUROSCIENCES GMBH.

XX Von Der Kammer H, Pohlner J;
XX WPI: 2003-845345/78.
XX N-PSDB; ACP58150.
XX
PT Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
XX
PS Claim 18; Fig 5; Opp; English.
XX
CC The invention relates to diagnosing or prognosticating a
CC neurodegenerative disease in a subject, or determining whether a subject
CC is at increased risk of developing the disease. The method involves
CC determining a level and/or an activity of a transcription or translation
CC product of a gene coding for golgin-245 and/or a fragment, derivative or
CC variant in a sample from the subject and comparing the level and/or the
CC activity to a reference value. The composition and methods are useful in
CC diagnosing, preventing and/or treating neurodegenerative diseases, such
CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
CC useful for screening, testing and validating compounds, agents and
CC modulators in the development of diagnostics and therapeutics to treat
CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
CC antibodies are used for detecting the pathological state of a cell in a
CC sample from a subject. The present sequence represents a human golgin-245
CC splice variant 1 polypeptide
XX
SQ Sequence 2228 AA;

Query Match 4.8%; Score 153.5; DB 7; Length 2228;
Best Local Similarity 19.7%; Pred. No. 0.0061;
Matches 135; Conservative 120; Mismatches 268; Indels 161; Gaps 31;

QY 13 NPSHDVYKVEKI-KIBSCSE-----RTKIFRLYSVCTVEEKALILARQNRKAVEN 64
DB 1552 NQOKDIEHKELVQKQHPQLGEBKDNVK-----EAEKILTLENQVYSMAKAELET 1603
QY 65 LNRSLTHALLMGA-----SYLFDKLDHFHSSSTPDGVSFQSGINDGVHIEPSSILSSK 119
DB 1604 KKKELEHVNLVSKSEELKALDELRESEAAKLAELKRAEQKIA-AIKQLLSQMEK 1662
QY 120 GGEENEVKLLLEAKHAQCTYSSDSTLFGEDEHKLSD-EESNIFWSKLLGKGNPMWKY 178
DB 1663 -----EEQYKKGTEHLSSEL-----NTKQEREREVHILEELKLSVESQ--- 1702
QY 179 PSDT---PQRNRKRVQYFEGSEASPTKGDGNNAKRRKASDDVTDPRVTDPPV-----D 229
DB 1703 -SETLIVPRSAKNVAAYTEQEEADSQ---GCYQKTYEEKISVLQRNLTEKEKLLQRVGQ 1757
QY 230 DIERKASGDKHMGALSPKVIITLQSSCKSGTDTLGG-----NDAPGLYSNGSHI 280
DB 1758 EKEETVSSHFMRCQYQOERIKLEHABAKQHEQDSMIGHLQEELEBKKNYSLT-VAQH 1816
QY 281 --SGIPEDMLASQDWKIPDESORRLHTVLKPKWAKLQVVLHS-----DACTSMVGNFLE 334
DB 1817 EREGKQNTQAKQNLNFDVDVQKTLQ-----EKELTCQILEQKIKELBSC-----LV 1864
QY 335 YVIENHRI-YEEPATTQAFQIALSWI-----AALLVKQILSHKGS 374
DB 1865 RQKEVHRVMEELTSKYELKQ-ALQQMDGRNKPTELEENTEESKSHLVQPKLLSNMEA 1923
QY 375 LVANSELAKRCSREVD-----YIYSILSCMSKSLFLEHTQGLQDCFGTNSKQSVST 428
DB 1924 ---QHNDLEFKLAGAREKQKJGKEIVRLQKDLRLRKEHQOELE-----ILKK 1969
QY 429 KLVNESLSGATVDEKINTK-----SMRNSDEECMTKRCSHYSTATRDIEKTIGIKK 484
DB 1970 EYDQE-----REEKIQOEDELKGNSTLQKLMREFN-TQLAQKEQELEWTIKETIN 2021
QY 485 KYKQVQKVLQVEHEEKKMELLNAYDKKQKLETSKSVEAAVIRITCSRSTQVGDULK--- 542

DB 2022 KAEVEAEELSHQBETNQLLKXIAEKODDLKRTAKRYEIEILDAREEEMTAKVRDLQTL 2081
QY 543 --LDHNYERKFEIKSEKNECLKSLEQMHDAVK-----KLADEACWINRIKSWAAK 593
DB 2082 BELQKYYOQKLEQENPGNDVNTIMELQTLQAOKTLLSDSKLKEQE--FRQIHNLDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
DB 2140 LKKYEKNVYATTGTPYKGNLYH 2163

RESULT 14
ABU07445
ID ABU07445 standard; protein; 2230 AA.
XX
AC ABU07445;
DT 28-JAN-2003 (first entry)
XX
DE Protein differentially regulated in prostate cancer #48.
XX
KW Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
OS Homo sapiens.
XX
PN WO200281638-A2.
XX
PD 17-OCT-2002.
XX
PF 08-APR-2002; 2002WO-US010824.
XX
PR 06-APR-2001; 2001US-0281731P.
PR 06-APR-2001; 2001US-0281732P.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Jay G;
XX
XX WPI; 2003-058520/05.
DR N-PSDB; ABX10347.
XX
PT Novel genes which are differentially regulated in prostate cancer, useful
PT for diagnosing prostate cancer in prostate tissue sample and assessing
PT therapeutic or preventive intervention in prostate cancer patients.
XX
PS Claim 1; Page 299-306; 416pp; English.
CC
CC The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) Is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) Is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity. (I)
CC is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type

of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways physiologically relevant to prostate cancer permits the definition of functional and disease pathways and the delineation of targets in these pathways which are useful in diagnostic, therapeutic, and clinical applications. This is the amino acid sequence of a protein differentially regulated in prostate cancer

DT	15-JAN-2004 (first entry)	
XX	Human golgin-245 splice variant 2 polypeptide.	
XX		
XX	Golgin-245; neurodegenerative disease; neuroprotective; neurotropic; antiparkinsonian; anticonvulsant; gene therapy; human; variant.	
KW		
XX	Homo sapiens.	
OS		
XX	WO2003087403-A2.	
PN		
XX	23-OCT-2003.	
PD		
XX	16-APR-2003; 2003WO-EP003958.	
XX		
PF	16-APR-2002; 2002EP-00008553.	
PR	16-APR-2002; 2002US-0372424P.	
XX		
XX	(EVOT-) EVOTEC NEUROSCIENCES GMBH.	
FA		
XX	Von Der Kammer H, Pohlner J;	
PI		
XX	WPI; 2003-845345/78.	
DR	N-PSDB; ACF58151.	
DR		
XX		
XX	Diagnosing or prognosticating a neurodegenerative disease comprises comparing a level or activity of a transcription or translation product of golgin-245 gene in a test sample with a reference value of a known disease or health status.	
PT		
PT		
PT		
XX	Claim 18; Fig 7; Opp; English.	
PS		
XX		
CC	The invention relates to diagnosing or prognosticating a neurodegenerative disease in a subject, or determining whether a subject is at increased risk of developing the disease. The method involves determining a level and/or an activity of a transcription or translation product of a gene coding for golgin-245 and/or a fragment, derivative or variant in a sample from the subject and comparing the level and/or the activity to a reference value. The composition and methods are useful in diagnosing, preventing and/or treating neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, or amyotrophic lateral sclerosis. The recombinant, non-human animal is useful for screening, testing and validating compounds, agents and modulators in the development of diagnostics and therapeutics to treat neurodegenerative diseases, in particular Alzheimer's disease. Specific antibodies are used for detecting the pathological state of a cell in a sample from a subject. The present sequence represents a human golgin-245 splice variant 2 polypeptide	
CC		
XX	Sequence 2230 AA;	
XX		
XX	Query Match 4.8%; Score 153.5; DB 7; Length 2230;	
XX	Best Local Similarity 19.7%; Pred. No. 0.0061;	
XX	Matches 135; Conservative 120; Mismatches 268; Indels 161; Gaps 31;	
Qy	13 NPSHDVKHVEKI-KIESCB-----RTKIFRLYSVCTVEEKALILARQNMRONKAVEN 64	
Db	1552 NQKQDIEHKLVKQLQHFQELGEEKNRVK-----EAEKILTLLENQVYSMKAELET 1603	
Qy	65 LNRSLTHALLMGA-----SYLPDKLDHFHSSTPDSGVSPFQISMDGVTHFSSILSSK 119	
Db	1604 KKKELEHVNLSVKSKBEELKALDRLESESAKLAELKKRAEQKIA-AIKKQLLSQMEEK 1662	
Qy	120 GGEENEVKLLCLEAKHAQCTYSDDSTLTFGEHDHKLSD-EESPNI FWSKLLGCKNPWKY 178	
Db	1663 -----EEQYKGTGSHLSEL-----NTKLQERERREHVHLEEKLLSVESQ--- 1702	
Qy	179 PSDT---PQRNKRKVQVFEGSEASPKTGCGNNAKKKASDDVTDPRVTPPV-----D 229	
Db	1703 -SETLIVPRAKNVAAYTEQEADSQ----GCYQKYEEKISVLQRNLTEKELKLQRVGQ 1757	
Qy	230 DDERKASGDKHMGALESPKVITILQSSCKSSGDTGTLDG-----NDAFGLYSMGSHI 280	

Db 1758 EKEETVSSHFERCQYQERLIKLEHAEAKOHEDEQSMIGHLOEBELEKKNKYSLI-VAQHV 1816
QY 281 --SGIPEDMLASQDMGKIPEQSORRLHTVLKPKMAKLCQVLHLS-----DACTSMVGNFLE 334
Db 1817 EKEGGKNWIOAKQNLNVDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
QY 335 YVIENHRI-YEBPATTFOAFQIALSWI-----AALLVKQILSHKES 374
Db 1865 RQKEVHRVEMEELTSKYKQLQ-ALQQMDGRNKPTLELENTBEKSKSHLVQPKLLSNMEA 1923
QY 375 LYRANSELAPKCSRVEVD-----YIYSILSCMKSLFLEHTQGLQDFDCTNSKQSVVST 428
Db 1924 ---QHNDLEFLAGAREKQKLGKEIVRLQKOLMLRKEHQOELE-----ILKK 1969
QY 429 KLVNESLSGATVRDEKINTK---SMRNSSEDEECWTEKRCSHYSTATRDIEKTIK 484
Db 1970 EYDQE-----REEKIQOEDELKXNSTLKQLMREFN-TQLAQKEQELEWTIKETIN 2021
QY 485 KYKQVQKLVQBHEEKKMELLNMYADKKQKLETSKSVEAAVIRITCSRTSTQVGDJLKL-- 542
Db 2022 KAQEVEAELLESQEBETNQLLAKIAEKDDDLKRTAKRYBEILDAREEEMTAKVRDLQTL 2081
QY 543 --LDHNYERKFDKSEKNECLKSLEQMDHVAKK-----KLADEACWINRIKSWAAK 593
Db 2082 BELQKKYQOKLEQENPGNDNVTIMELQTLAQKTTLISDSKLKEQE--FREQIHNLDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
Db 2140 LKYEKNVYATTVTGTPYKGNLYH 2163

Search completed: July 12, 2005, 00:58:34
Job time : 68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:11:48 ; Search time 16.6667 Seconds
(without alignments)
2781.424 Million cell updates/sec

Title: SEQ3_D1219_COPY_650_1270
Perfect score: 3217
Sequence: 1 RADAFILGSSLNPSHDVXH.....SGNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160.5	5.0	1739	4	US-09-976-594-76
2	160.5	5.0	1739	4	US-09-538-092-824
3	160.5	5.0	1740	4	US-09-949-016-8860
4	146	4.5	1939	4	US-09-949-016-6925
5	146	4.5	1942	4	US-09-949-016-8135
6	144.5	4.5	630	4	US-09-248-796A-20275
7	144.5	4.5	3210	4	US-09-538-092-1154
8	144.5	4.5	3248	1	US-08-353-700-1
9	144.5	4.5	3248	5	PCT-US95-16216-1
10	143.5	4.5	1960	4	US-09-538-092-1077
11	143.5	4.5	1960	4	US-09-949-016-10872
12	143.5	4.5	2482	1	US-08-328-254-6
13	142.5	4.4	1688	4	US-09-248-796A-14286
14	142.5	4.4	1972	4	US-08-875-435B-4
15	142	4.4	1836	4	US-09-949-016-7432
16	142	4.4	2733	4	US-09-949-016-11433
17	138	4.3	1695	4	US-09-866-108A-15753
18	137.5	4.3	1427	4	US-09-538-092-1044
19	135.5	4.2	8931	4	US-08-714-741-32
20	135	4.2	834	4	US-09-187-999-11
21	135	4.2	3259	4	US-09-949-016-6507
22	134.5	4.2	1857	4	US-09-917-254-91
23	134.5	4.2	1972	4	US-09-538-092-1084
24	134.5	4.2	1984	4	US-09-949-016-7111
25	134.5	4.2	1984	4	US-09-949-016-7112
26	134.5	4.2	1984	4	US-09-949-016-7113
27	133	4.1	1388	4	US-09-976-594-296

28	132	4.1	901	4	US-09-248-796A-14747	Sequence 14747, A
29	132	4.1	1388	2	US-08-685-576-1	Sequence 1, Appli
30	132	4.1	1786	4	US-09-949-016-7880	Sequence 7880, Ap
31	131.5	4.1	700	4	US-09-107-532A-5094	Sequence 5094, Ap
32	131.5	4.1	2186	4	US-09-949-016-10828	Sequence 10828, A
33	131.5	4.1	2349	4	US-09-538-092-914	Sequence 914, App
34	131	4.1	649	3	US-09-134-001C-3891	Sequence 3891, Ap
35	131	4.1	1976	4	US-09-538-092-1078	Sequence 1078, Ap
36	130.5	4.1	810	4	US-09-248-796A-20281	Sequence 20281, A
37	130.5	4.1	1401	4	US-09-750-590A-2	Sequence 2, Appli
38	130.5	4.1	1712	4	US-09-949-016-9450	Sequence 9450, Ap
39	130	4.0	1780	4	US-09-949-016-6859	Sequence 6859, Ap
40	130	4.0	1939	3	US-09-310-187A-1	Sequence 1, Appli
41	130	4.0	1939	4	US-09-538-092-917	Sequence 917, App
42	129.5	4.0	963	4	US-09-914-259-22	Sequence 22, Appl
43	129.5	4.0	1979	4	US-09-949-016-6468	Sequence 6468, Ap
44	129.5	4.0	2047	4	US-09-949-016-7404	Sequence 7404, Ap
45	129.5	4.0	2329	3	US-08-755-587-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-976-594-76
; Sequence 76, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1
US-09-976-594-76

Query Match	5.0%;	Score 160.5;	DB 4;	Length 1739;
Best Local Similarity	18.9%;	Pred. No. 2.5e-05;		
Matches 152;	Conservative 115;	Mismatches 260;	Indels 279;	Gaps 35;
QY	2	ADAFILGSSLNPSHDVXHVEIKIKIESCSERTKIFRLYSVCTVEKA-----L 49		
Db	879	ADTVIFDSDWPNQDLQARAHRIQKQKNIYKLVGTGVEEIIERAKKQVLDHL 938		
QY	50	ILARQMRQKAVEN-----LNRSLTHALLMWGASYLFDKLDHFSHSPDSCGVSF 100		
Db	939	VIQRMDDTTCRTILENNSSGRSNPNFKEILTALFKGASDLFKELE-----GEES 988		
QY	101	ESIMDGVTHFSSILSSKGGNEVKLLLEAKHAQGTYSDDSLTFGDHKLKLSDEES 160		
Db	989	EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQFKVANFATMEDEEL-----EER 1038		
QY	161	BNIFKSKLGGKPMWYFSDTPQRNRKVQYFE-----GSEASPKTG 203		
Db	1039	PHKMDDEII-----PEQRKKVEEERKELEIYMLPRISSSTKKAQTNDS 1085		
QY	204	DGNNAKKRKASDVDPVTPPPVDDDERKASGKDHMGALSPKVTILQSSCKSSGTDG 263		
Db	1086	DSDTESKQAQRSSASETED---SDDDKK-----PKRGRPRSVRKDLVEG 1130		
QY	264	TLDGN-----DAFLYSMGSHISGIPEDML--ASQWKGKIPESQRRHLTVLKPWAKLC 316		

```
Db 1131 FTDAEIRRFKAYKFF-----GLFLERLECIARD-AELVDKSVADL-----KRLG 1174
Qy 317 QVHLSDACTSMVGNFLEYVNIENHRIYEBPATTFOAFOIALSWIAALLVKQLSHKESLV 376
Db 1175 ELIH--NSCVSAMQYEEQLKENASGKPGKR-RGPTTIKISGV-QVNVKSIHQHEEPE 1230
Qy 377 RANSEL-----AFKCSRVEVDYIYSILSCWKSFLFHTQGLQDFCFGTNSK 422
Db 1231 MLHKSIPVDPEKKYCLTCRVKAAHFDVWGVEDDSRLLLGIYEHGYG-NWELIKTDPE 1289
Qy 423 QSVVSTKLVNESLSGATVDEKINTKSMNSSE-----DEECWTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLRKLEKKGAVTGGEAKLKKR 1342
Qy 465 -----CSHYSTATRDIEKTIISGIKKYKKYKQVQLVQHEEKK 501
Db 1343 KPRVKKENKVPRLKEEGIELSSPRSDNPSEEGVDGDLKSPMKKKQK-KKENKENK 1401
Qy 502 MELLNMYADK-----KQLETSSKVEA-----AVIRITCSRTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKDKKEPKSGDAKSSSKSRSGQPVHITAGSEVPVIGEDD 1461
Qy 540 -----LKLTDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQDQETFSICKERMRPVKALKQDKPKGLNVQGLEHTRNCLLKIGDRIAECLKAYS 1521
Qy 565 -----LEQMHVAKKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DQEHKLMWRNLWIFVSKTFEPDARKLHKLYMAHKRSQEEB-----EQKKDDV 1572
Qy 599 PIQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGKKPFRPEASGSRDLSISQS 1595

RESULT 2
US-09-538-092-824
; Sequence 824, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 824
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number O14647
US-09-538-092-824

Query Match 5.0%; Score 160.5; DB 4; Length 1739;
Best Local Similarity 18.9%; Pred. No. 2.5e-05;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

Qy 2 ADAPILGSSLSPSHDVKHVEKIKIESCSERTKIFLYSVCTVEEKA-----L 49
Db 879 ADTVVIFSDWNPQNDLOQARAHRIQOKKQNIYRLVTKGTVEEBIIBRAKKCMVLDHL 938
Qy 50 ILARQNMQRQKAVEN-----LNRSLTHALLMMGASVLPDKLDPHFSSTPPDSGVSF 100
Db 939 VLRQMDTTGRTILENNSGNSNPFNKBELTAILKFGAEDLFKELE-----GEES 988
```

```
Qy 101 EQSINDGVIHPESSILSKSGGEENVKCLLEAKHAQCTYSSDSTLFGEDHIKLSDEBS 160
Db 989 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQFKVANFATMEDEBEL-----EBR 1038
Qy 161 PNIPIFKLLGGKNPMWKYPSTPQRNRKRVQYFE-----GSEASPKTG 203
Db 1039 PHKOWDEII-----PEQRKKVEEERQKELEEIYMLPRISSTKKQATNDS 1085
Qy 204 DCGNAKKKASDDVDTPRVDPVDDDERKASGDHMGALSPKVIITLQSSCKSSGTGD 263
Db 1086 DSDTESKQOQRSSASESETD-----SDDDKK-----PKRGPRSVRKDLVEG 1130
Qy 264 TLIDGN-----DAFLYSMGSHISGIPEDML--ASQDMGKIPDESQRLLHTVLKPKMAKUC 316
Db 1131 FTDAEIRRFKAYKFF-----GLFLERLECIARD-AELVDKSVADL-----KRLG 1174
Qy 317 QVHLSDACTSMVGNFLEYVNIENHRIYEBPATTFOAFOIALSWIAALLVKQLSHKESLV 376
Db 1175 ELIH--NSCVSAMQYEEQLKENASGKPGKR-RGPTTIKISGV-QVNVKSIHQHEEPE 1230
Qy 377 RANSEL-----AFKCSRVEVDYIYSILSCWKSFLFHTQGLQDFCFGTNSK 422
Db 1231 MLHKSIPVDPEKKYCLTCRVKAAHFDVWGVEDDSRLLLGIYEHGYG-NWELIKTDPE 1289
Qy 423 QSVVSTKLVNESLSGATVDEKINTKSMNSSE-----DEECWTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLRKLEKKGAVTGGEAKLKKR 1342
Qy 465 -----CSHYSTATRDIEKTIISGIKKYKKYKQVQLVQHEEKK 501
Db 1343 KPRVKKENKVPRLKEEGIELSSPRSDNPSEEGVDGDLKSPMKKKQK-KKENKENK 1401
Qy 502 MELLNMYADK-----KQLETSSKVEA-----AVIRITCSRTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKDKKEPKSGDAKSSSKSRSGQPVHITAGSEVPVIGEDD 1461
Qy 540 -----LKLTDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQDQETFSICKERMRPVKALKQDKPKGLNVQGLEHTRNCLLKIGDRIAECLKAYS 1521
Qy 565 -----LEQMHVAKKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DQEHKLMWRNLWIFVSKTFEPDARKLHKLYMAHKRSQEEB-----EQKKDDV 1572
Qy 599 PIQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGKKPFRPEASGSRDLSISQS 1595
```

RESULT 3

```
US-09-949-016-8860
; Sequence 8860, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 8860
; LENGTH: 1740
; TYPE: PRT
; ORGANISM: Human
```

US-09-949-016-8860

Query Match 5.0%; Score 160.5; DB 4; Length 1740;
Best Local Similarity 18.9%; Pred. No. 2.5e-05;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY 2 ADAPILFGSLNPSHDVHVKHEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
DB 880 ADTVVIFDSWNPQNDLOAARHRIQOKKQVNIYRLVTKGTVEEBIERAKKMWLDHL 939
QY 50 ILARQNRQKAVEN-----LNRLTHALLMAGASYLPDKLDHFSSETPDSGVSF 100
DB 940 VIQRMDTTGTEILNENSGNSNPNFKEELTAILKFGAEDLFAELE-----GEES 989
QY 101 EQSIMDGVITHEFSSILSSKGGEEVVKLCILLEAKHAQGTYSDDSTLFGEDHIKLSDEES 160
DB 990 EPQEMD-----IDEILRLAETRENEVSTSTATDELL--SQFKVANFATMEDEEL-----EER 1039
QY 161 PNIFWKLGGKPMWKYPSDTQPNRKRVOYFE-----GSEASPKTG 203
DB 1040 PHKWDDEII-----PEQRKVEEERQKELEEIYMLPRIRSSTKCAQTND 1086
QY 204 DGGNAKRRKASDDVTDPVDDDERKASGDHMGALBSPKVTITLQSSCKSGTGG 263
DB 1087 DSUTESKQORQASASETED---SDDKK-----PKRRGRPRSRVKDLVEG 1131
QY 264 TLDPGN-----DAFLYSMGSHISGIPEDML--ASODWGKIPDESORRLHTVLKPKMAKLC 316
DB 1132 FTDAETRRFIKAYKF-----GLPLERLECIARD--AELVDKSVADL-----KRLG 1175
QY 317 QVHLSDACTSMVGNFLEVYVNIHRIYEPATTFQAFQALSIAALLVQILSHKESLV 376
DB 1176 ELIHT--NSCVSAMQVEEQLKENASEGKPGKR--RGPTIKISGV-QVNVKSIHQHEEPE 1231
QY 377 RANSEL-----AFKCSRVEDYIYVILSCMKSLFLEHTQGLQDFCFGNSK 422
DB 1232 MLHKSIPVDPPEKKYCLTCRVKAHFDVWEGVEDDSRLLLGIYEHGYG--NWEILIKTDPE 1290
QY 423 QSVVSTKLVSLSGATVDEKINTKSMRNSSE-----DEECMTEKR 464
DB 1291 -----LKLTDKILPVET--DKRQGLQTRADYLLKLRKGLKKGAVTGGEAKLKR 1343
QY 465 -----CSHYSTATRDIETKISGIIKKYKQVQKLVQEHHEKK 501
DB 1344 KPRVKKENVPRLKEBHEGIELSPRSDNPSEGEVKDDGLEKSPMKKKQK--KKENK 1402
QY 502 MELLNMYADK-----KQLETSSVEA-----AVIRITCSTSTQVGD--- 539
DB 1403 EKQMSRKOKEGDKERKKSKOKKPKSGDAKSSSKRSKSGPVHITAGSEPVPIGEDED 1462
QY 540 -----LKLDPH-----NYERKFD-----EIKSEKNECLKS-- 564
DB 1463 DDLDOETFTICKERMPVKALKQDKPKGLNVQOELEHTRNCLKIGDRTAECLKAYS 1522
QY 565 -----LEQMHVAKKGLAEDEACWINRIKSWAAKLKVCV 598
DB 1523 DQSHIKLWRRLNMFVSKFTEFDARKLHKLKYMAKKRSQEEB-----EQKKDDV 1573
QY 599 PQSGNNKH-----SCSSN-----ISON 617
DB 1574 ---TGGKCPFRPEASGSSRDSLISQS 1596

RESULT 4
US-09-949-016-6925
; Sequence 6925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6925
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6925

Query Match 4.5%; Score 146; DB 4; Length 1939;
Best Local Similarity 20.1%; Pred. No. 0.0007;
Matches 136; Conservative 105; Mismatches 256; Indels 180; Gaps 27;

QY 19 KHVEKIKIESCSER-TKIFRLYSVCTVEEKAIIILARQNRQKAVENLNRSLTHALLMWG 77
DB 972 KHATENKVNKLTEEMAGLDETIAKLTKEKKALQEAHQQTLDLQMEE----- 1018
QY 78 ASVLPDKLDHFSSETPDSGVSEFQSIMDGVITHEFSSILSSKGGEEVVKLCILLE--AK 135
DB 1019 -----DKVNTLTAKT-----KLEQQVDD-----LEGSLQEKKLCMDLERAKR 1057
QY 136 HAQG--TVSSDSTLFG--DHIKLSDEESPNIFWKLGGKPMWKYPSDTQPNRKRQV-- 191
DB 1058 KLEGDLKLAQESTMTENDKQQLNEKIKKKEFEMSNLOKQIIEDEQALANLQOKKIELQA 1117
QY 192 YFGESEASPKTDGGNAKRRKASD-----DVTD-----PRVTDPPVD--DDEKASGKD 239
DB 1118 RIEELEEEIEAERASRAKAKQKRSDLRELEISERLEBAGCATSAQIELNKKREAEFOK 1177
QY 240 HMGALBSPKVTITLQSSCKSGTGTGLDGNDAFLYSMGSHISGIP----- 284
DB 1178 MRDLE-----ESTLQHEATAAALRKKHADSVAELGQIDSLQVRVKQKLEKESELKM 1230
QY 285 -----EDMLASODWGKIPDESORRLHTVLKPKMAKL--- 315
DB 1231 EINDLASNMETVSKANFEKMCRTLEDQLSI---KTKEEQORLINELSAQKARLHTE 1287
QY 316 -----COVLHSDACTSMVGN-----FLEVYVNIHRIYEE-----PATTFQAFQALSIAAL 363
DB 1288 SGFESRQLDEKDAWYSQLSRGKQAFQOIEELKRLQLEETKAKSTLAHALQSA--RHDCDL 1346
QY 364 LVKQILSHKES-----LVTRANSALFKCSRVEDYIYVILSCMKSLFLEHTQGLQDFC 416
DB 1347 LREQYEEEOEAKAELQRGMSKANSEVAQWRTKYETDAI-----QRTTEELE--- 1391
QY 417 FGTNSKQSVVSTKLVSLSGATVDEKINTK-----SMRNSSEDECM-----TEKR 464
DB 1392 -----EAKKLAQRLODAEBEHVAVNSKASLEKTKQRLQNEVEDMLDIVERSNAA 1442
QY 465 CSHYSTATRDIETKISGIIKKYKQVQKLVQEHHEKK-----MELLNMYADKQKLETS 518
DB 1443 CTALDKQRNFDKVLAEWKQKVEETQAELEASQKESRSLSTELFKVKNAYEESLDHLETL 1502
QY 519 KSVEAAVIRITCSTSTQVGDILKLDHNYERKFDIKSEKNECLKSLEQMHVAKKGLAE 578
DB 1503 KR--ENKNLQOEISDLTEQIAEGGKTHLELEKVKQQLDHEKSELQTSLEE---AEASLEH 1557
QY 579 DEEA-----CWINRIKS 589
DB 1558 EBGKILRIQLELNQVKS 1574

RESULT 5
US-09-949-016-8135
; Sequence 8135, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapSeqFormat Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
; US-09-538-092-1154

Query Match 4.5%; Score 144.5; DB 4; Length 3210;
Best Local Similarity 21.0%; Pred. No. 0.0022;
Matches 150; Conservative 101; Mismatches 237; Indels 225; Gaps 37;

QY 7 LFSSSL-----NPSHDVHKVEKIKIESCSERTKIPFLYSVCTVEEKALILARQNM 57
Db 1622 VFCSLQEEENLTKRTPSAPAKGVE--ELESLE---VYR-QSLEKLEEK---MESQIM 1672

QY 58 QNKAVERNLRSLTHALLMWCASVLPDKLDFHFSSETPDSCVSPFQSIMDGVHIEFSSILS 117
Db 1673 KNKEIQELEQ-----LLSERQELCLRKQYLSENEQ-----WQKLTSTVLEMSKLA 1721

QY 118 SKGGEENEVKLCILLEAKHAQGYSSDSTLFG---EDHIKLSDEESPNFWSKLLGGKNP 174
Db 1722 AEKKQTEQLSLEVARLQGLDLSRSLGIDTDAIQ-GRNESCDI--SK-----1771

QY 175 MWKYPSPDPORNRKR-VQYFEGSEAS-----PKTGDGNAKRRKASDDVTDPRVTD 226
Db 1772 --EHTSETTTPKHVDVHOICDKDAQDQLNLDIEKITETGAVKPTGECSPDNTNYP 1829

QY 227 PVDDDERKAGKDHMGALSPKVTITLQSSCKSGTGTGLDGNDAFGLYSMGSHISGIPED 286
Db 1830 PGED---KTQG-----SSECLIS---ELSPGPNAL-----VPM 1857

QY 287 MLASQD-----WGKIPDESORRLHTV-----LKPMAKLCQVILHLS-----322
Db 1858 FLGNOEDIHNLQRLVKETSNENLRLLHVIDRDRKVBESLLNEMKELDSKLHLQEVOLMTK 1917

QY 323 -DACTSM---VGNP-----LEVNIENRIYEPATTFQAFQIALSW-----359
Db 1918 IEACIELEKTVGELKKNESDLSEKLYFCDHQELLQORVETSEGLNSDLEMHADKSSRED 1977

QY 360 IAALLVKOILSHKESLVRANSALA-PKCSRVVDYIYVILSCMKSLFLE-----HTOGL 412
Db 1978 IGVNAKVDSWKERFLDVNENLSRIRSEKASIEH-----EALYLEADLEVVQTEKL 2029

QY 413 QPFCFGTNSQSVVSTKLNVNLSGATVR-----TE-----KRSCHYSTATRDIEK---TISGI 482
Db 2030 CLEKDNENKQVIV---CLEELSIVTTSERNQLRGELDTWSKKTALTALDQLSEKMKETQE 2086

QY 452 NSEDEECM-----KRSCHYSTATRDIEK---TISGI 482
Db 2087 LESHQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQELQSLKDSQALSLT 2146

QY 483 KKKYKKQVQKLVQEH---EKKMELNNYADKK--QKLETSKVEAAVIR-----ITCSRT 533
Db 2147 KCELENOIAQLNKEKELLLVKESESLOARLSSESDEYKLVNLSKALEAALVKEGEPALRUS 2206

QY 534 STQVGLKLLDHNHYERKFEIK-----SEKN-----ECLKSLQMDHDAVAKKL 576
Db 2207 QEEV-----HQLRRGIEKLRVRIEADKQKHIAEKLKERERENDSLKDKV 2252

RESULT 8
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 4.5%; Score 144.5; DB 1; Length 3248;
Best Local Similarity 21.0%; Pred. No. 0.0022;
Matches 150; Conservative 101; Mismatches 237; Indels 225; Gaps 37;

QY 7 LFSSSL-----NPSHDVHKVEKIKIESCSERTKIPFLYSVCTVEEKALILARQNM 57
Db 1622 VFCSLQEEENLTKRTPSAPAKGVE--ELESLE---VYR-QSLEKLEEK---MESQIM 1672

QY 58 QNKAVERNLRSLTHALLMWCASVLPDKLDFHFSSETPDSCVSPFQSIMDGVHIEFSSILS 117
Db 1673 KNKEIQELEQ-----LLSERQELCLRKQYLSENEQ-----WQKLTSTVLEMSKLA 1721

QY 118 SKGGEENEVKLCILLEAKHAQGYSSDSTLFG---EDHIKLSDEESPNFWSKLLGGKNP 174
Db 1722 AEKKQTEQLSLEVARLQGLDLSRSLGIDTDAIQ-GRNESCDI--SK-----1771

QY 175 MWKYPSPDPORNRKR-VQYFEGSEAS-----PKTGDGNAKRRKASDDVTDPRVTD 226
Db 1772 --EHTSETTTPKHVDVHOICDKDAQDQLNLDIEKITETGAVKPTGECSPDNTNYP 1829

QY 227 PVDDDERKAGKDHMGALSPKVTITLQSSCKSGTGTGLDGNDAFGLYSMGSHISGIPED 286

```
Db 1830 PGED-----KTOG-----SSBCIS-----ELSFSGPNAL-----VPM 1857
Qy 287 MLASOD-----WGKIPDESORRLHTV-----LKPQMAKLCQVLHLS----- 322
Db 1858 FLGNQEDIHNLQRLVKETSNENLRLLHVIEDRDRKVESLLNEMKELDSKLHLEQVQLMTK 1917
Qy 323 -DACTSM-----VGNF-----LEVVIENHRIYEEBPATTFQAFQIALSW----- 359
Db 1918 IEACIELEKIVGELKKNESDLSEKLEFYFSCDHOELLQRVETSEGLNSDLEMHADKSSRED 1977
Qy 360 IAALLVKQILSHKESLVRANSOLA-FKCSRVEVDYIYLSLCKMSLFILE-----HTQGL 412
Db 1978 IGDNVAKVNDMSKRFDLVENLSRIRSEKASIEH-----EALYLEADLEVVQTEK 2029
Qy 413 QPDCFGTNSKQSVSTKLVNESLSGATVR-----DEKINTKSMR 451
Db 2030 CLEKDNENKQKIV---CLEEELSUVTSERNQLRGELDTMSKTTALDQSEKMEKTOE 2086
Qy 452 NSSEDEECM-----TE-----KRCSHYSTATRDIEK-----TISGI 482
Db 2087 LESHQSECLHLCIQVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALSIT 2146
Qy 483 KKKYKQVQKLVQEHF--EKKWELLMYADKK--OKLETSKSVEAAVIR-----ITCSRT 533
Db 2147 KCELENQIAQLNKEKELLVKESESLQARLSSESDYKLVNSKALEAALVEKGFALRLSST 2206
Qy 534 STQVGDGLKLLDHNRYERKFDIEK-----SEKN-----ECLKSLEQMHVAKKKL 576
Db 2207 QEEV-----HQLRGIEKLVRVIEADEKKQLHIAEKLKERERENDSLKDKV 2252
```

```
RESULT 9
PCT-US95-16216-1
; Sequence 1. Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet B.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
```

```
; ANTI-SENSE: NO
PCT-US95-16216-1
Query Match 4.5%; Score 144.5; DB 5; Length 3248;
Best Local Similarity 21.0%; Pred. No. 0.0022;
Matches 150; Conservative 101; Mismatches 237; Indels 225; Gaps 37;
Qy 7 LFGSSL-----NPSHDVHKVHKIKIESCSERTKIFRLYSVCTVEEKALILARQNNR 57
Db 1622 VFCSSLOEENLRKTRKTPSAPAKGVE--ELESLSCE--VVR-QSLEKLEBK--MESQIGIM 1672
Qy 58 QNKAIVENLRSLTHALLMWGASLYLFDKLDHFHSSSTPPSGVSGFQSIMDGVHIESSILS 117
Db 1673 KNKEIQELEQ-----LLSSERQELDLCKRQVLSNEQ-----WQOKLTSVTLEMSKLA 1721
Qy 118 SKGGGENEVKLLLEAKHAQGYSSDSTLFG---EDHKLSDEESPNIWFMSKLLGGKNP 174
Db 1722 AEKQTEQLSLEVARLQGLDLSSRSLLGIDTDAIQ-GRNESCII--SK----- 1771
Qy 175 MWKYPSTPQNRKR-VQYFEGSEAS-----PKTGDGNAKKRKKASDDVTDPRVTD 226
Db 1772 --EHTSETTERTPKHDVHQICDKDAQDLNLDIEKITETGAVKPTGECGSEQSPDNTVEP 1829
Qy 227 PVDDEKASGDHMGALLESKPVIITLQSCSKSGTDTGLDGNDAFGLYSMGSHISGIPED 286
Db 1830 PGED-----KTOG-----SSBCIS-----ELSFSGPNAL-----VPM 1857
Qy 287 MLASOD-----WGKIPDESORRLHTV-----LKPQMAKLCQVLHLS----- 322
Db 1858 FLGNQEDIHNLQRLVKETSNENLRLLHVIEDRDRKVESLLNEMKELDSKLHLEQVQLMTK 1917
Qy 323 -DACTSM-----VGNF-----LEVVIENHRIYEEBPATTFQAFQIALSW----- 359
Db 1918 IEACIELEKIVGELKKNESDLSEKLEFYFSCDHOELLQRVETSEGLNSDLEMHADKSSRED 1977
Qy 360 IAALLVKQILSHKESLVRANSOLA-FKCSRVEVDYIYLSLCKMSLFILE-----HTQGL 412
Db 1978 IGDNVAKVNDMSKRFDLVENLSRIRSEKASIEH-----EALYLEADLEVVQTEK 2029
Qy 413 QPDCFGTNSKQSVSTKLVNESLSGATVR-----DEKINTKSMR 451
Db 2030 CLEKDNENKQKIV---CLEEELSUVTSERNQLRGELDTMSKTTALDQSEKMEKTOE 2086
Qy 452 NSSEDEECM-----TE-----KRCSHYSTATRDIEK-----TISGI 482
Db 2087 LESHQSECLHLCIQVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLSLEKDSQALSIT 2146
Qy 483 KKKYKQVQKLVQEHF--EKKWELLMYADKK--OKLETSKSVEAAVIR-----ITCSRT 533
Db 2147 KCELENQIAQLNKEKELLVKESESLQARLSSESDYKLVNSKALEAALVEKGFALRLSST 2206
Qy 534 STQVGDGLKLLDHNRYERKFDIEK-----SEKN-----ECLKSLEQMHVAKKKL 576
Db 2207 QEEV-----HQLRGIEKLVRVIEADEKKQLHIAEKLKERERENDSLKDKV 2252
```

```
RESULT 10
US-09-538-092-1077
; Sequence 1077. Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
```

; SEQ ID NO 1077
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35579
US-09-538-092-1077

Query Match 4.5%; Score 143.5; DB 4; Length 1960;
Best Local Similarity 19.4%; Pred. No. 0.0012;
Matches 126; Conservative 91; Mismatches 207; Indels 227; Gaps 23;
QY 22 EKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQRKAVENLNRLSLTHALLMWSGASYL 81
Db 960 QKLEKVTTEAKLKKL-----EEQIILEDQNCCLAK----- 992
QY 82 FDKLDHFSSETPDSGVSPFQSDMGVIEHFSSILSKGGEENEVKLLLEAKHAQGTY 141
Db 993 -----EKKLEDRIAEFTTNLT-----EEEKSKSLAKLKNKH----- 1025
QY 142 SSDSTLFGEDHDKLSDEESNIFWSKLLGGKNPMWKYPSTPORNKRKVOYFEGSEASP 201
Db 1026 -----EAMITDLE-----ERLRREKQROELE----- 1047
QY 202 TGDGNAKRRKASDDVTPRVDDDERKASGDHMGALSPKVITLQSSCKSSGT 261
Db 1048 -----KTRKLEGSTD-----LSDQIAELQAOIAELKQWLAKKEE-----ELQALARVEE 1094
QY 262 DGTLDGNDAPGLYSMGSHIGIPEDMLASQDWGKIPDESORRLHTVLKPKMAKLCQVLHL 321
Db 1095 EAAQKNMALKKIRELESQISELQEDLESERASRNKAEKQKRDGLGELEALKTELEDTLDS 1154
QY 322 SDACTSMVGNFLEYVIEHRIYEEPTTFQAFQI-----ALSWIAALL----- 364
Db 1155 TAAQQLRSKRGQEVNLIKLTLEEEAKTHEA-QIQEMRQKHSQAVELEAQLBQTKRVKA 1213
QY 365 -----VKQIL-----SHKESLV-----RANSELAF 384
Db 1214 NLEKAKQTLNERGELANEVKVLQKGDSEHKKVKEAQQLQVKNFNEGERVETLAD 1273
QY 385 KCSR--VEVDYIYSILSCMSLF-----LEHTQGLQDFCGTNSKQSVVSTK 429
Db 1274 KVTQLQVELDNVTGLLSQSDSKSLTKDFSALESQLODTQELLQE---ENRQKLSLSTK 1330
QY 430 LVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEK-----TI 479
Db 1331 L-----KQVEDEKNSFREQLSEEEBAKHNEKQIATLHAQVADMKKQWEDSVGCLETA 1383
QY 480 SGIKKKYKQVQKLVQEHKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 539
Db 1384 EYVKRLQKDLGLSQPHEEK-----VAAY-DKLEKTKT-----RLQQLDLD 1424
QY 540 LKL-LDH-----NYERKFEIKSEKNECLKSLEQMHDAKCKLADE 580
Db 1425 LLVDLDHORQACNLEKKQKFPDQLAEKTIKSAKYAEERDRAEAREKE 1475

RESULT 11

US-09-949-016-10872
; Sequence 10872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10872
; LENGTH: 1960
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10872

Query Match 4.5%; Score 143.5; DB 4; Length 1960;
Best Local Similarity 19.4%; Pred. No. 0.0012;
Matches 126; Conservative 91; Mismatches 207; Indels 227; Gaps 23;
QY 22 EKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQRKAVENLNRLSLTHALLMWSGASYL 81
Db 960 QKLEKVTTEAKLKKL-----EEQIILEDQNCCLAK----- 992
QY 82 FDKLDHFSSETPDSGVSPFQSDMGVIEHFSSILSKGGEENEVKLLLEAKHAQGTY 141
Db 993 -----EKKLEDRIAEFTTNLT-----EEEKSKSLAKLKNKH----- 1025
QY 142 SSDSTLFGEDHDKLSDEESNIFWSKLLGGKNPMWKYPSTPORNKRKVOYFEGSEASP 201
Db 1026 -----EAMITDLE-----ERLRREKQROELE----- 1047
QY 202 TGDGNAKRRKASDDVTPRVDDDERKASGDHMGALSPKVITLQSSCKSSGT 261
Db 1048 -----KTRKLEGSTD-----LSDQIAELQAOIAELKQWLAKKEE-----ELQALARVEE 1094
QY 262 DGTLDGNDAPGLYSMGSHIGIPEDMLASQDWGKIPDESORRLHTVLKPKMAKLCQVLHL 321
Db 1095 EAAQKNMALKKIRELESQISELQEDLESERASRNKAEKQKRDGLGELEALKTELEDTLDS 1154
QY 322 SDACTSMVGNFLEYVIEHRIYEEPTTFQAFQI-----ALSWIAALL----- 364
Db 1155 TAAQQLRSKRGQEVNLIKLTLEEEAKTHEA-QIQEMRQKHSQAVELEAQLBQTKRVKA 1213
QY 365 -----VKQIL-----SHKESLV-----RANSELAF 384
Db 1214 NLEKAKQTLNERGELANEVKVLQKGDSEHKKVKEAQQLQVKNFNEGERVETLAD 1273
QY 385 KCSR--VEVDYIYSILSCMSLF-----LEHTQGLQDFCGTNSKQSVVSTK 429
Db 1274 KVTQLQVELDNVTGLLSQSDSKSLTKDFSALESQLODTQELLQE---ENRQKLSLSTK 1330
QY 430 LVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEK-----TI 479
Db 1331 L-----KQVEDEKNSFREQLSEEEBAKHNEKQIATLHAQVADMKKQWEDSVGCLETA 1383
QY 480 SGIKKKYKQVQKLVQEHKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 539
Db 1384 EYVKRLQKDLGLSQPHEEK-----VAAY-DKLEKTKT-----RLQQLDLD 1424
QY 540 LKL-LDH-----NYERKFEIKSEKNECLKSLEQMHDAKCKLADE 580
Db 1425 LLVDLDHORQACNLEKKQKFPDQLAEKTIKSAKYAEERDRAEAREKE 1475

RESULT 12

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego

```
/ STATE: California
/ COUNTRY: USA
/ ZIP: 92122
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/328,254
/ FILING DATE: 24-OCT-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/141,239
/ FILING DATE: 22-OCT-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Campbell, Cathryn A.
/ REGISTRATION NUMBER: 31,815
/ REFERENCE/DOCKET NUMBER: P-CJ 1191
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 535-9001
/ TELEFAX: (619) 535-8949
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2482 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-328-254-6

Query Match 4.5%; Score 143.5; DB 1; Length 2482;
Best Local Similarity 21.0%; Pred. No. 0.0018;
Matches 150; Conservative 101; Mismatches 237; Indels 225; Gaps 37;

QY 7 LFGSSII-----NPSHDVHVEKIKIESCERTKIERLYSVCTVSEKALILARQNR 57
Db 894 VFCSSIQEENLRTKTPSPAPAGVE--ELESICE---VYR-QSELEKLEEK---MESQGM 944
QY 58 QNKAVENLNRLTHALLMWCASYLFDKLDHFSSTPDSGVSEFQIMDGVIEHFSIILS 117
Db 945 KNKEIQELSEQ-----LLSSERQELDLKQYLSENEQ-----WQKLTSVTLEMESKLA 993
QY 118 SKGGEENEVKLLLEAKHQAQTYSSDSTLFG---EDHKLDESSEPNITFWSKLLGGKNP 174
Db 994 ABEKKQTEQLSLEVARLQGLDLSSRLGIDTDEDAIQ-GRNESCDI--SK-----1043
QY 175 MWKYPSTDPQRKR--VQYFEGSEAS-----PKTGDGNAKRRKKASDDVTDPRVTD 226
Db 1044 --EHTSETTERTPKHUVHQCCKDAQODLNLDIEKITETGALKPTGECGEGSPDTNYPE 1101
QY 227 PVDDDERKASGDHMGALSPKVTITLQSSCKSGTDLGNDADFGLYSMGSHISGIPED 286
Db 1102 PGED---KTQG-----SSCIS---ELSPGPNAL-----VPM 1129
QY 287 MLASQD-----WQKIPDESORRLHTV-----LKPMAKLCQVILHS-----322
Db 1130 FLNGQEDIHNLQLRVKETSNEENLRLHVEDRDKVESLLENKELDSLHLQEVQLMTK 1189
QY 323 -DACTSM---VGNF-----LEVLIENHRIYEPPATTQAFQIALSW-----359
Db 1190 IBACIELEKIVGELKENSDDLSEKLEYFCSDHQELLQRVETSEGLNDSLEMIADKSSRED 1249
QY 360 IAALLVQKILSHKESIVRANSELA-FKCSRVEVDITYIISLCMKSIFLE-----HTQGL 412
Db 1250 IGDVAKVNDWMKERFLDVENELSRIRSEKASIEH-----EALYLEADLEVQVTEKL 1301
QY 413 QPDCFGTNSKQSVVSTKLVNESLSGATVR-----DEKINTKSMR 451
Db 1302 CLEKONENKQKIV---CLEELS VVTSERNQLRGELDTWMSKKTALTALDQSEKMKETQOE 1358
QY 452 NSEDEECM-----TE-----KRCSHYSTATRIEK---TISGI 482
Db 1359 LESHQSECLHCIOVAEAEVKEKTELLQTLSSDVSELLKOKTHLOEKLQSEKDSQALSILT 1418
```

```
QY 483 KKKYKKQVQKLVQEHHE--EKKMELLMNYADKK--OKLETSKSVAEAVIR-----ITCSRT 533
Db 1419 KCELENQIAQLNKEKXELLVGESESIQARLSSESDYEKLNYSKALEAALVEKGFALRLSST 1478
QY 534 STQVGDKLKLLDHNHYERKPEDEIK-----SEKN-----ECLKSLEQMHVDVAKKL 576
Db 1479 QEEV-----HQLRRGIEKLRVRIEADKKQLHTAEKLRERERENDSLKDKV 1524

RESULT 13
US-09-248-796A-14286
; Sequence 14286, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14286
; LENGTH: 1648
; TYPE: PRT
; ORGANISM: Candida albicans
/ US-09-248-796A-14286

Query Match 4.4%; Score 142.5; DB 4; Length 1648;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 122; Conservative 100; Mismatches 227; Indels 155; Gaps 26;

QY 60 KAVENLNRLTHALLMWCASYLFDKLDH-----FHSSETPDSGVSEFQIMDGVIEH- 112
Db 900 KNLERLNLSLKKIKELWQTLFYIQTNPIYLTFLFNSIPYTKNQTSGODLFSQVILPFP 959
QY 113 ---SSILSKSGGEENEVKLC-LLEAKHQAQTYSSDST-----146
Db 960 VRDSIITHSREYFLVKMLQMLQMDNTANSNLDGITKHLTNWIDFTFNHNTFQ 1019
QY 147 ----LFGEDHKLSD-----EESNIFWSKLLGGKPNMWKYPSTDPQRNRKVQYFEGS 196
Db 1020 HLKALLGKFIKRVINEQVDPESDPIRYNQII--DHEMKVYGRSEKSRDISPQAAIQLP 1077
QY 197 EASPTGDDGNAKRRKKASDDVTDPRVTDPPVDDDERKASGK-----DHM-----CA 243
Db 1078 EYSNKF--VGNLMSURETCSDLLSML-----QKASGNKALLQIPDHVKLICROQY 1126
QY 244 LESPKVITLQSSCKSGTDLGNDADFGLYSMGSHISGIPED---MLASQDWGKIPD-E 299
Db 1127 LCAQRKFPDKSQOQHLAVAGVI-----FVKHVLGS-ILQVPENYGLTGNNDTQAKSKD 1180
QY 300 SORRLHTV-----LKPMAKLCQVILHSDACTSMVGNFLYEVJENHRIYE- 344
Db 1181 NLRLYRVMLQLFSMPKPNDFNLKPLNE---YIMASTDTVKSIIISQAIINVGEIETVVEL 1237
QY 345 ---EPAITFOAFQIALSWIAL-LVKQILSHKESIVRANSELA-FKCSRVEVDITYIISLSC 400
Db 1238 HDYDDLVTQRKPLTISVNSLQLEKSIQNDVIITGNDQDLQYK-TCVEVEKL--LISP 1294
QY 401 MKSLFLEHTQGLQFCFGTNSKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECM 460
Db 1295 QDMLTLDLSSVTNLNPTQESIVDSK-----TKTLFQA-----1331
QY 461 TEKRCSHYSTATRD-----IEKTSIGIKKKYKQVQKLVQEHHEKQMLLMNYADKKQ 513
Db 1332 --KRCLLYIIRVQEEEDSDDLLELLISGIKPSHEQRFKEIVQYEK-----AEQDI 1379
QY 514 KLETSKSVAEAVI---RITCSRTSTQVGDKLKLLDHNHYERKPEDEIKSEKNECLKSLEQMH 570
```


Db 1380 SLNKSQVNTVANKKSTRPYSGTSLGDSNLTYH-----ELKWCLEIILKESMGE 1433
QY 571 VAKK 574
Db 1434 LTRK 1437
RESULT 14
US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; PRIOR FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4
Query Match 4.4%; Score 142.5; DB 4; Length 1972;
Best Local Similarity 18.3%; Pred. No. 0.0015;
Matches 117; Conservative 108; Mismatches 212; Indels 201; Gaps 25;
QY 22 EKIKESCERTKIFRLYSVCTVEEKAIIILARQNMQRKAVENLNRLTHALLMGCASYL 81
Db 967 QKLEKVTAEAKIKKL-----EDDILVMDQNNKLSK----- 999
QY 82 FDKLDHFHSETPDGVSFQSGTMDGVIHEFSSILSSKGGEEVNEVKLLLEAKHAQGY 141
Db 1000 -----ERKLEERISDLITNLAE-----EEKAQNLTKLNKH----- 1032
QY 142 SSDSTLFGEDHIKLSDEESNIFWSKLLGKNPMWKYPSTPQ-----RNR 187
Db 1033 ---ESMISELEVLKKEERSQLEKL---KRMDGEASDLHQIADLQAIAELKWLQA 1086
QY 188 KRVOYFEGSEA--SPKTDGCKNAKGRKASDDVTPRVTDPVVD--DDEKASQKDHMGAL 244
Db 1087 KKEELQAALARLEDETQSNNALKKIRE---LEGHISLDQLDLSERAARNK---AE 1138
QY 245 ESPKVTITQSSCKSGTDTGLDGNDAFLGYSMSGSHISGIPEDMLASQDNGKIDFESQRL 304
Db 1139 KQRDLGEELEAKTELEDTLD-----TTATQQLERAKRE-----QE 1175
QY 305 HTVLKPRM-----AKLCQVLHLSDACTSMVGNFLEYVNIENHRIYEBPATTFFQAFIALSW 359
Db 1176 VTVLKALDETRSHAEQVQEMQKHTQVVEELTEQLEQPKRANLDTKQTKLEKENAD 1235
QY 360 IAALL-----VKQILSHKESLVRAN--SELAFKCS-----RVEV-DVIYSILSCMKSLFLE 407
Db 1236 LAGELRVLQAKQVEVHKKKLQVQLQELQSKSGDGERARAEIENDKVHKLQNEVESV--- 1292
QY 408 HTQGLQFDCGNTSKQSVSTKLVNESLS--GATVRD-----EKNKTSMRNSSED 456
Db 1293 -----TGMUSEAEGKAIKLAKVASIGSLQDQLQELLQELBETQKLNVTSLKRLQLED 1343

QY 457 BECWTEKRCRSHVSTATRDIEKTIISI-----KKYKKQVQK 492
Db 1344 ERNSLOQLDEEMEAQNLERHISTLNIOQLSDSKKQLQDFASTVESLEEGKGFQKEIES 1403
QY 493 LJQHEHEKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDGLKL-LDH----- 545
Db 1404 LTOQYEEK-----AAAYDKLETK-----NRLOQELDDLVDLQNRQLVS 1444
QY 546 --NYERKPFDEIKSEKNECKLSLEQHWVAKKKLAED 580
Db 1445 NLEKKQKFDQLLAEKNISSKYADERDRAEAAREKE 1482
RESULT 15
US-09-949-016-7432
; Sequence 7432, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7432
; LENGTH: 1836
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7432
Query Match 4.4%; Score 142; DB 4; Length 1836;
Best Local Similarity 19.7%; Pred. No. 0.0015;
Matches 125; Conservative 98; Mismatches 217; Indels 194; Gaps 29;
QY 2 ADAPILFGSLNPSHDVKEVKEIKESCERTKIFRLYSVCTVEEKA-----L 49
Db 822 ADTVIFDSQNPQNDLQQAARHIGQKKQVNIYRLVTGTVVEEIIIRAKKKVLDHL 881
QY 50 ILARQNMQRKAVEN-----LNRLTHALLMGCASYLFDKLDHFHSETPDGVSF 100
Db 882 VIQRMDTTGRTILENNGSRNSNPNKELTALIKFGAEDLPKLE-----GEES 931
QY 101 EQSIMDGVITHEFSSILSSKGGEEVNEVKLLLEAKHAQGYSDSTLFGEDHIKLSDEES 160
Db 932 EQEMD-----IDEILRLAETRENEVSTSATDELL--SQFKVANFATMEDEEL-----EER 981
QY 161 PNIFWSKLLGKNPMWKYPSTDPQRNRKRVQVFESEASPKTDGCKNAKGRKASDDVTD 220
Db 982 PHKMDDEII-----PEQRKKVE-----EEERQKLEBEI-- 1010
QY 221 PRVTPPVDDEKASQKDHMGALSPKVTITQSSCKSGTDTGLDGNDAFLGYSMSGSHI 280
Db 1011 --YMLPRISSTKKAQTNDSDSDESKR-----QAQRSSASETESDD----- 1053
QY 281 SGIPEDMLASQDNGKIDFESQRLHHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYVNIENH 340
Db 1054 -----DKKPKRG---RPRSVKRLDVEGTDA-----EIR 1080
QY 341 RIYEEPATTFQAFIALSWIAALLVKQILSH-----KSLVRANSELAPK----- 385
Db 1081 RLVEA-----LFSLSLFSAYFAQLSQAQTNDSDSDESKRQAQRSSASETES 1132
QY 386 -----CSR-VEVDVIYSILSCMK-----LFEHTQ--GL---QFPCFGTNSKQSVV 426

```
Db 1133 DDDKKPRKRCPRSVKDPRLRSL-CMDTDLHMFIKAYKKFGLPLERLECIARDA----- 1186
QY 427 STKLVNESLSGATVRDEKINTKSMRNSSEDEECMTEKRCGSHYSTATR---DIEKTISGIK 483
Db 1187 --ELVDKSVADLKRGLGELIHNCSVSAWQYEEBOLKENASEGKPGKRGPTIK--ISGV- 1241
QY 484 KKYKKQVQKLVQHEHEEKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVG---DL 540
Db 1242 ---QVNVKSIHQHEEFEMLHKSI PVDPEKKK-----YCLTCRVKAAHFDVWGVEDDS 1293
QY 541 KLL-----DHYERKPFDEIKSEKNECLKSLQMHID 570
Db 1294 RLLLGIEHGY-GNWELIKTDPELKLTDKEITHD 1326
```

Search completed: July 12, 2005, 01:04:35
Job time : 20.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:58:28 ; Search time 57.6667 Seconds
(without alignments)
4160.324 Million cell updates/sec

Title: SEQ3_D1219_COPY_650_1270
Perfect score: 3217
Sequence: 1 RADAFILFGSSLNPSHDVHKH.....SGNNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1726218 seqs, 386331768 residues

Total number of hits satisfying chosen parameters: 1726218

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3211	99.8	2254	17 US-10-732-923-18943
2	590.5	18.4	1990	16 US-10-437-963-178706
3	495.5	15.4	2630	16 US-10-437-963-172374
4	456.5	14.2	1954	16 US-10-437-963-178698
5	335.5	10.4	358	16 US-10-425-115-343695
6	192.5	6.0	162	15 US-10-424-599-155808
7	163.5	5.1	1461	17 US-10-732-923-8689
8	163.5	5.1	1465	15 US-10-369-493-5100
9	163.5	5.1	1465	17 US-10-732-923-8690
10	160.5	5.0	1738	17 US-10-732-923-8335
11	160.5	5.0	1739	15 US-10-341-434-61

12	160.5	5.0	1739	17 US-10-732-923-8832	Sequence 8832, Ap
13	159	4.9	1445	17 US-10-732-923-8160	Sequence 8160, Ap
14	157	4.9	126	16 US-10-425-115-329771	Sequence 329771, Ap
15	157	4.9	1015	15 US-10-369-493-3210	Sequence 3210, Ap
16	154.5	4.8	1358	15 US-10-675-072A-43	Sequence 43, Appl
17	151	4.7	2552	16 US-10-437-963-129734	Sequence 129734, Ap
18	150	4.7	1786	17 US-10-732-923-8672	Sequence 8672, Ap
19	149.5	4.6	699	16 US-10-425-115-293739	Sequence 293739, Ap
20	149	4.6	1881	14 US-10-032-585-7646	Sequence 7646, Ap
21	147.5	4.6	728	17 US-10-732-923-8570	Sequence 8570, Ap
22	146	4.5	1959	15 US-10-028-248A-36	Sequence 36, Appl
23	146	4.5	1959	15 US-10-107-782-35	Sequence 35, Appl
24	145.5	4.5	1813	17 US-10-732-923-8155	Sequence 8155, Ap
25	144.5	4.5	3114	16 US-10-751-736-99	Sequence 99, Appl
26	143.5	4.5	1960	15 US-10-236-031B-62	Sequence 62, Appl
27	143.5	4.5	1960	15 US-10-028-248A-104	Sequence 104, App
28	143.5	4.5	1960	15 US-10-107-782-104	Sequence 104, App
29	143.5	4.5	1960	17 US-10-696-909A-49	Sequence 49, Appl
30	143.5	4.5	3113	16 US-10-723-860-862	Sequence 862, App
31	142.5	4.4	1422	17 US-10-732-923-8665	Sequence 8665, Ap
32	142.5	4.4	1647	14 US-10-032-585-7572	Sequence 7572, Ap
33	142.5	4.4	6761	17 US-10-732-923-15035	Sequence 15035, A
34	142	4.4	1379	14 US-10-205-219-5	Sequence 5, Appl1
35	142	4.4	1426	17 US-10-732-923-8310	Sequence 8310, Ap
36	142	4.4	3225	16 US-10-408-765A-254	Sequence 254, App
37	141	4.4	865	17 US-10-732-923-7041	Sequence 7041, Ap
38	141	4.4	1961	15 US-10-028-248A-103	Sequence 103, App
39	141	4.4	1961	15 US-10-107-782-103	Sequence 103, App
40	140.5	4.4	1643	17 US-10-732-923-8217	Sequence 8217, Ap
41	140.5	4.4	1806	17 US-10-732-923-8571	Sequence 8571, Ap
42	140	4.4	1130	15 US-10-369-493-6751	Sequence 6751, Ap
43	139.5	4.3	1829	17 US-10-732-923-8687	Sequence 8687, Ap
44	138.5	4.3	760	16 US-10-437-963-188325	Sequence 188325, Ap
45	138	4.3	972	9 US-09-924-154-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-10-732-923-18943
; Sequence 18943, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18943
; LENGTH: 2254
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-18943

Query Match 99.8%; Score 3211; DB 17; Length 2254;
Best Local Similarity 99.8%; Pred. No. 9.9e-238;
Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	RADAFILFGSSLNPSHDVHKHVKIESCSERTKIFRLYSVCTVEEKALILARQNQRK 60
DB	883	RADAFILFGSSLNPSHDVHKHVKIESCSERTKIFRLYSVCTVEEKALILARQNQRK 942
QY	61	AVENLNRSITHALLMGASYLFDKLDHFSSTPDGVSFEQIMDGVITHEFSSILSSKG 120
DB	943	AVENLNRSITHALLMGASYLFDKLDHFSSTPDGVSFEQIMDGVITHEFSSILSSKG 1002
QY	121	GEENEVKLCILLEAKHAQCTYSSDSTLFGEDHKLSDDESPNIFWSKLGGKPNMKYPS 180
DB	1003	GEENEVKLCILLEAKHAQCTYSSDSTLFGEDHKLSDDESPNIFWSKLGGKPNMKYPS 1062

bad date

Db 881 DAIILYSDNNPLNDLKALQKIESQFERVSIFRLYPTFVEKSLVLARQGVIDNNI 940
Qy 63 ENLNRSLTHALLMWGASYLDKLDHSHSSSTPDSGVSEFQSDIMDGVTHFSSILSSKGE 122
Db 941 QDLRTSLKSLRWGAFLSRDLVEQDDHASKSEMERHFDIVELTKLSTTVED 1000
Qy 123 ENEVKLCLEAKHAQTYSSDSTLFGEDHILKLSDESPNIFWSKLLGGKXNPMWYPSDT 182
Db 1001 STEVHRKSIKANNMGBGLSIRNITLMEKEGIVLEDNPAEFLWNLDDGRSPHYSCISEP 1060
Qy 183 PQRNRKEVOYFEGSEASPKTGDGNAKRRKASDDV---TDPRTDPPVDD---DERKAS 236
Db 1061 LQSRVTKSQTMD---EVNAPAEINEARKKHKVGEIMGSSKVSQSDKSDNDALPDICTTS 1118
Qy 237 GKDHMGALESPPKVTILQSSCKSSGTDLGDNDAFGLYSMSHISGIPEDMLASQDWGKI 296
Db 1119 GP---ALQ-PVDVTQKSVQSEGESLM-----1142
Qy 297 PDSQRRLHTVLKPKMAKLCQVHLHS-----DACTSMVGNFLEXVIE 338
Db 1143 --STPKNLHAQMKQELSCLKVLQLPNTIQNGAKIAGQVLIIEVDNVTLLVEQFFEYLLN 1200
Qy 339 NRIYIEPATFOAFQIALSWIAALLVKILSHKESLVRANSELAFKCSRVEVDYIYSIL 398
Db 1201 NHVVQEPKTYFHALNIALCWRVASIHNFKVDHKSALAEKRLKYECNEELARLVYDSL 1260
Qy 399 SCMKSLFLEHTQGLQFCFGTNSKQSVSTKLVNESLSGATVRDEKINTK-----448
Db 1261 --KRPEKAGATGNCQST---SVEKTKPQOQETSN-ILANDHIFPKQRMDLHDNFM 1312
Qy 449 -----SMRNSSEDECM---TEKRCSHYSTATRD--TEKTSIGIKKKYKQVQK 492
Db 1313 NGALQEGSFVAAQWVSEQLIAVPGTHMEC-HFSTDELDPDIVEKRIINLIDNVFSREYR 1371
Qy 493 LVOEHEEKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFD 552
Db 1372 IFDQSQSISELEKYQNTKARLKT-----VCN-----LVLEHICRSHAD 1411
Qy 553 EIKSEKNECLSLQEHMDHVAKKLADEACWINRIKS-----WA-----AKLKVCVPTQSG 603
Db 1412 --VETRNDTIKQVWFQWFTMLMYAFLEHMLHLSKLESLSQNTWAEERQLKELCLEAKSG 1469
Qy 604 -----NNKHFGSSS 612
Db 1470 QLDHTFDQIALPDSNFMQEFTHLKEQSSNSHVSGSA 1507

RESULT 4

US-10-437-963-178698
; Sequence 178698, Application US/10437963
; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 178698

; LENGTH: 1954

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1.pap

US-10-437-963-178698

Query Match 14.2%; Score 456.5; DB 16; Length 1954;

Best Local Similarity 24.0%; Pred. No. 1.6e-25;

Matches 160; Conservative 118; Mismatches 229; Indels 159; Gaps 24;

Qy 3 DAFILFGSSLNPHSDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNRQKAV 62

Db 755 DAIILYCSMDNPTDLVLRQISIESQSECVPIFRLYSSCTVEKTLIL-----803

Qy 63 ENLNRSLTHALLMWGASYL---FDKLDHSHSSSTPDSGVSEFQSDIMDGVTHFSSILSSK 119

Db 804 ----RSM--IIFATSRILEEFQKID-YSSKQSEDDGL-----FMNVFLFAAKLSTN 851

Qy 120 GGENEVKLCLEAKHAQTYSSDSTLFGEDHILKLSDESPNIFWSKLLGGKXNPMW 176

Db 852 VEASTKMNNAVIGRAQSGSFYSRDIAVISEREGISAVDGLPKFWTFWNLGGRSPHW 911

Qy 177 KYPSDTPQRNRKEVOYFEGSEASPKTGDGNAKRRKASDDVTPRTDPPVDDDERKAS 236

Db 912 QYISEPVQRNRKIQNMEDQMRIPABETDEAIMKRKIGE-----IMD-----954

Qy 237 GKDHMGALESPPKVTILQSSCKSSGTDLGDNDAFGLYSMSHISGIPEDMLASOD---292

Db 955 -----SPKILPVKD-----NDV-----LPENSTASSSHETS 982

Qy 293 ---WGKIPDES---QRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEXVIEHRIYE 344

Db 983 VDDTWQELGAESLQGTQKGLHTQKPELSKLYELLELPETVKLCBELLDVILKNHQ---1039

Qy 345 EPATTQAFQIALSWIAALLVKILSHKESLVRANSELAFKCSRVEVDYIYSILSKSL 404

Db 1040 -----CMRAASLLKHKNRRESLALSVRLNAYECDEVLAEYVYKLRILKCK 1086

Qy 405 FLEHT-----QGLQFPCFGTNSKQSVSTKLVNES--LSGATVRDEKINTKSMRNSSE-456

Db 1087 FSRRASETSKQSQSTPVNTSSYKQTSKPLRSDGSGICHQVTTIDGLENVSHHEAPHDI 1146

Qy 457 ---EECTEKRCSHYSTATRDIEKTSIGIKKKYKQV--QKLVOEHEEKMELNMYADKK 512

Db 1147 LTEMILEQK-----ELISVLETHREHVLRLDELLERITEKRINLINMVFSLR 1194

Qy 513 QK-----LETSSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFDIEKSEKNE 560

Db 1195 EKNIQDKQGNETTLLDMHKKQKQVAKLQETCN-----LVVEHLRKGHID--SEDRDA 1243

Qy 561 CLKSLEQMDHVAKKLADEACWINRIK-----SWAAKLKVC-VPIQSGNKHFGSSSNI 614

Db 1244 TVKLITIEWFTLLLYAFINHMRCQHKNLKMQQSTSWNKELQKELFLOQAAGSHLDRSFDQ 1303

Qy 615 SQNAPD 620

Db 1304 QIPLPD 1309

RESULT 5

US-10-425-115-343695

; Sequence 343695, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 343695

; LENGTH: 358

; TYPE: PRT

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(358)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_76611C.1.pap
US-10-425-343695

Query Match
Best Local Similarity 10.4%; Score 335.5; DB 16; Length 358;
Matches 102; Conservative 58; Mismatches 120; Indels 49; Gaps 12;

QY 3 DAFILGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQRKAV 62
DB 47 DAIILYCDWNPMDLRLQRVSMESQSPVPFIRLYSFTVEEKALILAKHDHILDSNI 106
QY 63 ENLNRSLTHALLMWGASYLFDKLDHP--HSSETPDSGVSEFQSGIMDGVHEFSSLSK 120
DB 107 VNITPSLSCHLLSWGASPLFNLEELQOHSY----SNVSGDELFMNDVDFLTKLLSKV 162
QY 121 GEENEVKLCLEAKHAQGYSSDSTLFGD-DHIKLSDESPNI--FWSKLLGKNPMWK 177
DB 163 ERTSGNTAISOAYLCGSFYRAIVVAGERGIPSDVDLKFWAYWVSLNGRSPQWQ 222
QY 178 YPSDTPQNRKRQVYFEGSEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASG 237
DB 223 YISEPVQSRRKI---NNMEQQLKNTD---KQLKITTEDEARV-----KRRIG 267
QY 238 K-DHMGALSPKVTITLQSSCKSGTGTLDGNDAGLYSMGSHISGIPEDMLASQDWGKI 296
DB 268 EIMSSAIDSP-----GKNKDTILFGNNT-----PPSSH-----QISVEDTWQEL 307
QY 297 P---DESQRRLHTVLKPKMAKLCOVHL 321
DB 308 ERSNLHATQGLHVQLKPEISKLYKLQL 336

RESULT 6
US-10-424-599-155808
; Sequence 155808, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 155808
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1.pap
US-10-424-599-155808

Query Match
Best Local Similarity 6.0%; Score 192.5; DB 15; Length 162;
Matches 57; Conservative 32; Mismatches 71; Indels 35; Gaps 4;

QY 337 IENHRIYEPATTFOAFQALSWIALLYKQILSHKESLVRANSELAFKSCRVVDYIYS 396
DB 1 MNLYNRRPFPSISQAFQSLCWTAAASLLKHLKLOPIASLIQ---DLNFECKEEDYIYS 57
QY 397 ILSCKMSLFEHTQGLQFCFGTNSQSVVSTKLNVNESLSGATVRDEKINTSMRNSSED 456
DB 58 MLCCLKKIFLYRTGNHYDTGF-----TKASGFSNRA 88
```

```
QY 457 BECTEKRCSHYSTATRDIEKTIISGIIKKYKQVOKLVQEHKEKQKELLNMYADKKQKLE 516
DB 89 YSCTGVAR--EVELFKOMSKSIKEIQKCEKKLKLHLQBEERQRLRAATEEBEKANFD 146
QY 517 TSKSVEAAVIRITCS 531
DB 147 ERYNIESADIR-SCS 160

RESULT 7
US-10-732-923-86689
; Sequence 8689, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8689
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-732-923-8689

Query Match
Best Local Similarity 5.1%; Score 163.5; DB 17; Length 1461;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADATILFSSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQRK 61
DB 799 ADTVIIFDSDWNPQNDLQAMSAHRIGTQKTVNIVRLVTKGSVEB-----BI 845
QY 62 VENLNRSLTHALLMWGASYLFDKLDH-----FHSSETPDSGVSEFQSGIMDGV 108
DB 846 VERAKRKLIV-----LDHLVIQRMDDTKTVLSKNATASGSPVDFKQELSAI 891
QY 109 IHERFSI--LSSKGGEENEV-----KLCLLLEAKHAQGYSSDSTLFGD-DHIKLSDES 160
DB 892 L-KFGAVELFKEGEGBEPEVDIDRILMGATRAEBEVMKENELSSFKYANFAIDEE 950
QY 161 PNIF-----WSKLLGKNPMWKYPDSPTQNRKR------Q 191
DB 951 KDIAAATDEWAAII-----PEEDRNRILEERMEKELAEMLNLAQRKQPIQ 997
QY 192 YFEGSEA-SPKTDGNGNAKRRKASDDVTDPRV-----TDPVDDDERK 234
DB 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRIFKFRKFSPLNRLIEIAQDAELE 1057
QY 235 ASGKDHMGALSPKVTITLQSSCKSGTGTLDGNDAGLYSMGSHISGIPEDMLASQDWG 294
DB 1058 EHSDEMCKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
QY 295 KIPDESQ-----RRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEV 336
DB 1093 AAESEKKDIERKFKPHTCDVNLKQIERSHAELKP-----LHEILKGEETKTS----- 1139
QY 337 IENHRIYEPATT-----FOAFQAL-----SWIAALIVKQI-LSHKESL 375
DB 1140 -----FKPPANAKLQKQWMDVMSRPDSSALLGLGWKVGKYGYSWEAIKMDPTGLADKIFI 1193
QY 376 VRANSELAFKSCRVVDYIYSILSCMSKSLFLEHTQGLQFCFGTNSQSVVSTKLNVNESL 435
DB 1194 KDTKKKPGQKNLOVRVDYLLKLM-----SKDKVKVKTTE----- 1225
QY 436 SGATVRDEKINTKSMRNSSEDECTEKRCSHYSTATRDIEKTIISGIIKKYKQVOKLVQ 495
DB 1226 -----KERRRRKADDDVPVGPCKKRRHTNNVPQEGE-----KKCKK----- 1262
QY 496 EHEKKMELLNMYADKKQKLETSKSAVAVIRITSRSTSTQVGDLLDHNVERKPFDEIK 555
```

Db 1263 --EKNSSSLK---DQALLSIDKSLYGGALESSAKPFLECVKLCMPVHKYMKLKEAQ 1317
QY 556 SEKNELCKS-----LEQMDHVAKKKLABDEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKQADEAKYLRGLGDSFLENLETLIKKPKTNIRKWNLYWFLCKFTLRP 1371

RESULT 8
US-10-369-493-5100
; Sequence 5100, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5100
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5100

Query Match 5.1%; Score 163.5; DB 15; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0041;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAPILFGSSLNPSHDVKEKIKIESCSERTKIFRLYSVCTVEEKALILARQMNKA 61
Db 799 ADTVIFDSWNPQNDQAMSAHRIGQTKTVNIYRLVTGSGVEE-----EI 845

QY 62 VENLNRLTHALLMWCASYLFDKLDH-----PHSSETPDSGVSPSEQIMDGV 108
Db 846 VERAKRLV-----LDHLVQRMDDTKTVLSKNATASGVPPDKQELSAI 891

QY 109 IHEFSSI--LSSKGGEBENEV-----KLCILLLEAKHAQTYSSDSTLFG--DHIKLSDERS 160
Db 892 L-KFGAVELFKEGESEQPEVDIDRLMCAETREAEVEEVMKENELLSFKYANFAIDEE 950

QY 161 PNIF-----WSKLLGKNPMWKYPSDTPQNRKRV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDRNRIIEERMKELAEMLNAPRQKQPIQ 997

QY 192 YFEGSEA-SPKTDGNGNAKKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRPIKFRKFSMPLNRLLEIAQDAELE 1057

QY 235 ASGKDHMGALSPKVTITLOSSCKSGTDLGNDAFGLYSMGSHISGIPEDMLASQDWG 294
Db 1058 EHSTDEMCKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092

QY 295 KIPDSQ-----RRLHTVLKPKMAKLCQVHLHSDACTSMVGNFLEYV 336
Db 1093 AAESKKDTERKPKFHTCDVNLKQIERSHAELKP-----LHEILKSBETKTS-----1139

QY 337 IENHRIYEPATT-----FOAFOIAL-----SWIAALLVKQI--LSHKSLS 375
Db 1140 -----FKPPANAKLQKQWVDWNSRPDDSGALLGVWYKYGSGWEAIKMPDTGLADKIFI 1193

QY 376 VRANSELAPKCSREVDYIYSILSCMKSLFLEHTQGLQFCFCFTNSQSWSTKLVNESL 435
Db 1194 KDKTKKPGQKNQVRVDYLLKLM-----SKDKVKTE-----1225

QY 436 SGATVRDEKINTKSMNSSEDEECMTKRCSHYSTATRDIEKTIISGIIKKYKQVQKLVQ 495

Db 1226 -----KKERKRKADDPVPGPEKKKRHTNNVPQSGEK-----KKKEKK-----1262
QY 496 EHEEKOMELNNYADKKQKLETSKVSVEAAVIRITCSRTSTQVGDGLKLLOHNYRERKFDK 555
Db 1263 --EKNSSSLK---DQALLSIDKSLYGGALESSAKPFLECVKLCMPVHKYMKLKEAQ 1317

QY 556 SEKNELCKS-----LEQMDHVAKKKLABDEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKQADEAKYLRGLGDSFLENLETLIKKPKTNIRKWNLYWFLCKFTLRP 1371

RESULT 9
US-10-732-923-8690
; Sequence 8690, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8690
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-732-923-8690

Query Match 5.1%; Score 163.5; DB 17; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0041;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAPILFGSSLNPSHDVKEKIKIESCSERTKIFRLYSVCTVEEKALILARQMNKA 61
Db 799 ADTVIFDSWNPQNDQAMSAHRIGQTKTVNIYRLVTGSGVEE-----EI 845

QY 62 VENLNRLTHALLMWCASYLFDKLDH-----PHSSETPDSGVSPSEQIMDGV 108
Db 846 VERAKRLV-----LDHLVQRMDDTKTVLSKNATASGVPPDKQELSAI 891

QY 109 IHEFSSI--LSSKGGEBENEV-----KLCILLLEAKHAQTYSSDSTLFG--DHIKLSDERS 160
Db 892 L-KFGAVELFKEGESEQPEVDIDRLMCAETREAEVEEVMKENELLSFKYANFAIDEE 950

QY 161 PNIF-----WSKLLGKNPMWKYPSDTPQNRKRV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDRNRIIEERMKELAEMLNAPRQKQPIQ 997

QY 192 YFEGSEA-SPKTDGNGNAKKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRPIKFRKFSMPLNRLLEIAQDAELE 1057

QY 235 ASGKDHMGALSPKVTITLOSSCKSGTDLGNDAFGLYSMGSHISGIPEDMLASQDWG 294
Db 1058 EHSTDEMCKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092

QY 295 KIPDSQ-----RRLHTVLKPKMAKLCQVHLHSDACTSMVGNFLEYV 336
Db 1093 AAESKKDTERKPKFHTCDVNLKQIERSHAELKP-----LHEILKSBETKTS-----1139

QY 337 IENHRIYEPATT-----FOAFOIAL-----SWIAALLVKQI--LSHKSLS 375
Db 1140 -----FKPPANAKLQKQWVDWNSRPDDSGALLGVWYKYGSGWEAIKMPDTGLADKIFI 1193

QY 376 VRANSELAPKCSREVDYIYSILSCMKSLFLEHTQGLQFCFCFTNSQSWSTKLVNESL 435
Db 1194 KDKTKKPGQKNQVRVDYLLKLM-----SKDKVKTE-----1225

QY 436 SGATVRDEKINTKSMNSSEDEECMTKRCSHYSTATRDIEKTIISGIIKKYKQVQKLVQ 495

```
Db 1226 -----KKERKRKADDDVPVGPPEKKKRHTNNVPQGEK-----KKKEKK----- 1262
QY 496 EHEEKMELNNYADKKQKLETSKSVAAAVIRITCSRTSTQVGLDKLLDHNHYERKFEIK 555
Db 1263 ---EKNSSSLK---DQALLSIDKSLYGGALEDSAKPFLCVCVKMPVHKYMKLKEAQ 1317
QY 556 SKNECLKS-----LEQMDVAKKGLADEACWINRIKSWAAKUKVCVP 599
Db 1318 EAKQADEAKYLTRLGDSFLENLETLIKKPKTNIRKWTNLIWFLCKFTLREP 1371

RESULT 10
US-10-732-923-8335
; Sequence 8335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8335
; LENGTH: 1738
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-8335

Query Match 5.0%; Score 160.5; DB 17; Length 1738;
Best Local Similarity 18.5%; Pred. No. 0.0089;
Matches 138; Conservative 104; Mismatches 234; Indels 269; Gaps 31;

QY 2 ADATILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNM----- 56
Db 1041 ADTVVIFSDWNPNQDLQAMRAHRIGQQEVNVIYFVTSKVEEILERAQKMWLDHL 1100
QY 57 -----RONKAVENLNRLTHALLMWGASYLPDKLDHFSHSETPSDGSVFEQS 103
Db 1101 VTQKLNAGREKRETKGSGNFDKNELSAIRFGAEELPKE----- 1141
QY 104 IMDGVIHEFSSILSSKGGENEVKLCLL-----LEAKHAQGTYSSTLFGEDH 152
Db 1142 -----DKNDESKRLLSMDITELERAEOVEEKH-----TDTEHELLGAFK 1184
QY 153 I-KLSEESPNIWFKLLGGKPNWKYPSTDPQNRKRQVYFEGSEASPKTGDGNAKR 211
Db 1185 VANFCNAEDDGGFSWRWKPDVSVTAEALAPRAARTKSYVDPSHPD-----RTSKRK 1238
QY 212 KKASDDVTPRVTPDPPVDDDERKASGDHMGALSPKVITLQSSCKSGTGTGLGNDAP 271
Db 1239 KKGSE-----PPEHTERSQKRKTEYFVPSTPLLEGTSQAQVGHSGYNLPKRDQA 1288
QY 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKI PDESQRRLHTVLKPKMAKLCQV 318
Db 1289 RFVRTVMKFGNHQMACIAEVEGVVE-----APEEAQVELFDA----- 1328
QY 319 LHLSDACTSMV--GNF-----LEYVTIENHRIYEPAATTQ 351
Db 1329 --LIDGCKESVETGNFEPKGPVLDFPGVPVKANELLKRVQGLQLSKRISRYNDPISQFR 1386
QY 352 AFQIALSWTA-----ALLVKQILSH---KESLVRANSELAF---KCSRVEV 391
Db 1387 -----VLSYKLPNSWKGCGNQIDDAELLGILYHGFNGWEKIRLDESGLTKKIAPVEL 1442
QY 392 DVIYISILCMKSLFLEHTQGLQDFC-----GTNSKQSVVSTKLVNESLSG-----ATVRDE 443
Db 1443 QHHTETFLPAPNLKERATALLEMELAAAGGKNTAKASRKNKVKVDNLINQKAPARDR 1502
QY 444 K-----INTK-----SMRNSSE-----DSE 458
Db 1503 RGKSGPANVSLSTKDGPRKTKAEPFLVKEGEMSDDGEVYEQFKSQKMWEWCEDVLADE 1562
```

```
QY 459 CMTEKRCRSHYSTATRDI--EKTISGIKKYK--KQVOKLYOEHEE-----KKMELLN 506
Db 1563 IKTLRLQRLQTTISADLPKPKVLFKIRYLEILGRRIDAIIVLEHEEDLYKQDRMTMLWN 1622
QY 507 -----MYADKKQKLETSKSVAAAVIRITCSRTSTQVGLDKLLDHNHYERKF 551
Db 1623 YVSTFNSLGBRLNQIYSKJQKEEBEGVGS--HLNGSR-----NFQRQ- 1666
QY 552 DEIKSEKNECLKSLEQMH---DVAK 573
Db 1667 QKFKTAGNS--QGSQOVHKGIDTAK 1689

RESULT 11
US-10-341-434-61
; Sequence 61, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-61

Query Match 5.0%; Score 160.5; DB 15; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0089;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY 2 ADATILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVVIFSDWNPNQDLQARAHRIQKQKQVNIYRLVTGTVEEIEIERAKKMWLDHL 938
QY 50 ILARQNMQRNKAVERN-----LNRLTHALLMWGASYLPDKLDHFSHSETPSGVSF 100
Db 939 VIQRMDTTGRTILENNSGRSNPNFKEELTALIKFGAEDLPKELE-----GEES 988
QY 101 EOSIMDGVIEHFPSSILSSKGGENEVKLCLLEAKHAQGTYSSTLFGEDHILKLSDEES 160
Db 989 EPQEMD-----IDEILRLAETRENEVSTSATDELL--SQPKVANFATMEDEEL---BER 1038
QY 161 PNIWFKLLGGKPNWKYPSTDPQNRKRQVYFE-----GSEASPKTG 203
Db 1039 PHQMDIELI-----PEQRKKVVEEERQKEELIYMLPRISSTTKKAQTND 1085
QY 204 DGGNAKRKKAQDDVTPRVTPDPPVDDDERKASGDHMGALSPKVITLQSSCKSGTGTG 263
Db 1086 DSDTESKRAQSSASESETD---SDDDKK-----PKRGRPRSVRKDLVVEG 1130
QY 264 TLJDN-----DAFGLYSGWSHISGIPEDML--ASQDWGKIPDESQRRLHTVLKPKMAKLC 316
Db 1131 FTDAEIRRFKAYKXF-----GLPLERLECLARD--AELVDKSVADL-----KRUG 1174
QY 317 QVYLHLSDACTSMVGNFLEVYIENHRIYEPAATTFOAQFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQSYEBEOLKENASGKPGKR--RGPTIKISGV--QVNVKSLIQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEVDYIYISILCMKSLFLEHTQGLQDFCFOTNSK 422
Db 1231 MLHKSIPVDPEKKKYCLTCRVKAAHFDVWGVEDDRLLLGLIYEHGYG--NMELIKTDPE 1289
QY 423 QSVVSTKLVNESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
```



```

Db      1290  -----LKLTKILPVET--DKPQKQLOTRADYLLKLLRKGLEKKGAVTGGSEAKLKKR 1342
QY      465  -----CASHYATRDIEKTIISGIIKKYKKQVQKLVQEHHEKK 501
Db      1343  KPRVKENKVPRLKEHGIELSSPRSDNPSEGEVKKDGLSKSPMKKKQK-KKENK 1401
QY      502  MELLNNYADK-----KQKLETSKVEA-----AVIRITCSTRTSTQVGD--- 539
Db      1402  EKQMSRKQKEGDKERKKSKDKEKPKSGDAKSSKSKSGPVPVHITAGSEPVPIGEDED 1461
QY      540  -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db      1462  DLDQETFSICKERMVRPVKALKQDKPDKGLNVQBLEHTRNCLLKIGDRIAECLKAYS 1521
QY      565  -----LEQMDHVAKKKLAEDACWINRIKSWAAKLKVCV 598
Db      1522  DQEHILKLRNLWIFVSKTFEFDARKLHLYKWAHKKRSQEEB-----EOKKKDDV 1572
QY      599  PQSGNNKHFF-----SGSSN-----ISQN 617
Db      1573  ---TGGKKPFRPEASGSSRDSLSQS 1595

```

RESULT 12

```

US-10-732-923-8832
; Sequence 8832, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-732-923-8832

```

```

Query Match      5.0%; Score 160.5; DB 17; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0089;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;

QY      2 ADAPILFGSLNPSHDVHKVEKIKIESCERTKIFRLYSVCTVEEKA-----L 49
Db      879  ADTVIFDSWNPQNDLOAQAARAHGQKKQVNIYRLVTKGTVEEBEIERAKKMWLDHL 938
QY      50  ILARONMRQKAVEN-----LNRLTHALLMWGASVLFDKLDHFHSSETPDGSGVGF 100
Db      939  VIQMDTTGTLIENNSGRSNPNFKBELTALKFGAEDLFKELE-----GEES 988
QY      101  EQSMDGVITHEFSSILSSKSGEENEVKLCLELLEAKHAQGYTSSDSTLFGEDHKLSDERS 160
Db      989  EPQEMD-----IDEILRLAETRENEVSTSATDELL--SQFKVANFATMEDEEL-----EER 1038
QY      161  PNIWFKLLGGKPMWPKYSDTPQNRKRVQYFE-----GSEASPKTG 203
Db      1039  PHKWDDEII-----PEQRKKVEEBERQKELEIYMLPRIRSTTKCAQTND 1085
QY      204  DGGNAKRRKASDDVTDPRVTPDPPVDDDERKASGDHMGALSPKVITLQSSCKSGTGG 263
Db      1086  DSDTESKQARSSASESETED---SDDKK-----PKRGRPRSRVKDLYEG 1130
QY      264  TLQGN-----DAFGLYSMGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLKPKNAKLC 316
Db      1131  FTDAETRRFFIKAYKFF-----GLPLERLECLARD-AEIVDKSVADL-----KRLG 1174
QY      317  QVHLHSDACTSMVGNFLEVYVNIENHRIYEBPATTFOAFQALSWIAALLVKQLSHKESLV 376

```

```

Db      1175  ELIH--NSCVSAMQYEEBOLKENASEGKPGKR-RGPTIKISGV-QVNVKSIIOHEEPE 1230
QY      377  RANSEL-----AFKCSRVEYDIYISILSCMKSLFLEHTQGLQDFCGTNSK 422
Db      1231  MLHKSIPVDPPEKKKYCLTCRKAHFDFVGEWVEDDSRLLLGIYEHYG-NWELIKTDPE 1289
QY      423  QSVVSTKLNVESLSGATVDEKINTKSMRNSB-----DRECHTEKR 464
Db      1290  -----LKLTKILPVET--DKPQKQLOTRADYLLKLLRKGLEKKGAVTGGSEAKLKKR 1342
QY      465  -----CASHYATRDIEKTIISGIIKKYKKQVQKLVQEHHEKK 501
Db      1343  KPRVKENKVPRLKEHGIELSSPRSDNPSEGEVKKDGLSKSPMKKKQK-KKENK 1401
QY      502  MELLNNYADK-----KQKLETSKVEA-----AVIRITCSTRTSTQVGD--- 539
Db      1402  EKQMSRKQKEGDKERKKSKDKEKPKSGDAKSSKSKSGPVPVHITAGSEPVPIGEDED 1461
QY      540  -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db      1462  DLDQETFSICKERMVRPVKALKQDKPDKGLNVQBLEHTRNCLLKIGDRIAECLKAYS 1521
QY      565  -----LEQMDHVAKKKLAEDACWINRIKSWAAKLKVCV 598
Db      1522  DQEHILKLRNLWIFVSKTFEFDARKLHLYKWAHKKRSQEEB-----EOKKKDDV 1572
QY      599  PQSGNNKHFF-----SGSSN-----ISQN 617
Db      1573  ---TGGKKPFRPEASGSSRDSLSQS 1595

```

RESULT 13

```

US-10-732-923-8160
; Sequence 8160, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8160
; LENGTH: 1445
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-732-923-8160

Query Match      4.9%; Score 159; DB 17; Length 1445;
Best Local Similarity 18.0%; Pred. No. 0.0089;
Matches 143; Conservative 112; Mismatches 212; Indels 328; Gaps 35;

QY      2 ADAPILFGSLNPSHDVHKVEKIKIESCERTKIFRLYSVCTVEEKA ILARONMRQ-- 58
Db      551  ANWVILFDPWNPANDLOAIDRAYRIGQCRD-VKVFRLISLGTVEE--IMYLRQVYKQOL 607
QY      59  -----NKAVENLRSLTHALLMWGASVLF-----DKLDH----- 87
Db      608  HCVVVGSENAKRYFAVQSGKEHGFVGNLFLKRLSGSGCLTRILREBQVEAGIMT 667
QY      88  -----FISSETPD-----SGVSFEQSIMDGVITHE----- 111
Db      668  ATTWLKGEASAKLETFRPDQCEPTNVNCELSDMSDEESVGVKAVGHVGGSRPTGSSA 727
QY      112  -----FSSILSK-----GGEENEVKLCLELLEAKHA--QGYTSSDSTL 147
Db      728  QLTLLQCGSKLFEAKCESVQDGNVPSSDGSDEQPMCLSAKARQAACQKTWDSVCT- 786
QY      148  FGEDHKLSDSESPNIFWSKLLGGKNPMWKYSDTPQNRKRVQYFEGSEASPKT----- 202
Db      787  --SEH-----QKSDNI-----QTPDEKCVSD---KSEKTLQNVSSSEDDTKCHST 828

```

```
QY 203 -----GDCGNKAKRKASDDVTDTP--RV-----TDPVPV-----DDDER 233
Db 829 AGHHCWGQDTESSDVFPTQDPQRPKNPICCKLLLGSESDSETPDKVKNHDDGR 888
QY 234 KASGK-----DHMGALSKPVITLQSSCKSSGTDGTLGNDADFGLYSMGSHISGIPEDML 288
Db 889 QNSGRGNPVSFLYLEN---MTLKSVRKRGKGTDDISDESDDIDMF-----PKSRI 936
QY 289 ASQDWG---KIPDESQRLLHTVLKPKWAK-LCOVLHLSDACTSMVGNFLEYVIENHRIYE 344
Db 937 RQQRATASLKLKSKKENKRNKHNPSITAKDTNQVCEADGDCSSQ-----VIEDPSSD 989
QY 345 EPATTFQAFQIALSWIAALLVKQI-----LSHKESLVRANSELAFKCSRVEVDYIYISLSC 400
Db 990 DSSS-----VSHVSFTKLHRAETVKDGVQ-----EVAYIH-----1020
QY 401 MKSLFLEHTQGLQFDCFGTNSKQSVVSTKLVNESLSGATVRD-----442
Db 1021 -----SNQNVIGSSRAENHMSRWATRDVPFLKQPSQLPANVAVCS 1060
QY 443 -----EKINTKSMRNSSEDE-----ECWTEKRCSHYST--ATRDIEKTISSG 481
Db 1061 STYKSKVNANAVIPTKQOPSPDGHISSPLVISHVPQSKKQKDVTRNTHTTIIIGETPRG 1120
QY 482 IKKK-----YKKQVKLVQVE-----HEKKM-----ELLN 506
Db 1121 IRRKQPEEMASFYKLSVKEFAEQVTSATSEEQRMRLRDFVSLQHPBKEFFPMDASLKM 1180
QY 507 MYADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNHYERKFDIEKSEKNECLKSLE 566
Db 1181 SVHEKEERVNRKSGKESLLKERPSNDSTL-----SCNDSTN 1217
QY 567 QMHDVAKKLADEA 581
Db 1218 KMSQVYNQKICEGKS 1232

RESULT 14
US-10-425-115-329771
; Sequence 329771, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 329771
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1.pap
US-10-425-115-329771
```

```
Query Match 4.9%; Score 157; DB 16; Length 126;
Best Local Similarity 35.4%; Pred. No. 0.00039;
Matches 40; Conservative 18; Mismatches 47; Indels 8; Gaps 3;

QY 283 IPEDMLASQDWGKIPD---ESQR-----RLHTVLKPKWAKLCOVLHLSDACTSMVGNFLEY 335
Db 10 LPDMSTPSADLHLPLPTGVENMRTPKSSFAELKRELSKLTNVLKLPDNLVFLANQSLLEY 69
QY 336 VTENHRIYEPATTFQAFQIALSWIAALLVKQI-LSHKESLVRANSELAFKCS 387
Db 70 LNNHLVVRPRSIHLAFNIALCWRAASFLLKTYELDRHRESLALASDGLAYECN 122
```

```
RESULT 15
US-10-369-493-3210
; Sequence 3210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiaofeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3210
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3210
```

```
Query Match 4.9%; Score 157; DB 15; Length 1015;
Best Local Similarity 20.1%; Pred. No. 0.0077;
Matches 129; Conservative 108; Mismatches 254; Indels 150; Gaps 28;

QY 17 DVKHVEKIKISCSERTKIFRLYSVCTVBEKALILARQNMQRQKAVENLRSLTHALLMW 76
Db 225 DIEHWRQVQTDLKKEDSLSL-----OEEA---TRKTEQADQMSLNRTLNL-----269
QY 77 GASYLDFDKLDHFSSETPDGVSPEQSIMDGVIEHFSSILSSKGGEENEVKLCLELEAK- 135
Db 270 -----DMDTAALGVPIKGADPHDIGATILEVFDNLNLNY--EDTEQRLVQVEAKD 321
QY 136 HAQGYTSSDSTLFGEDHI-KLSDEES-----PNIPWSKLLGGKNPMWYPSDT 182
Db 322 KAAAAAKESATRSRSETIQLKTSNNKLKEQDEQKADAFYQKL---KKKGWLEPTED 378
QY 183 P-----ORNRKRVQYFEGSEASPKTGDGNGNAKR-KKASDDVTDPRV---TTPPVDD 230
Db 379 PSASIALDQKSLKQDQDRWESEKSNALNOKITNLNKSIEBANVALKMDKLMQVERDLSVDL 438
QY 231 DERKASGKDHMGALSKS---PKVITLQSSCKSSGTDGTLGNDADFGLYSMGSHISGIPEDML 288
Db 439 QQRQE-----GDIKSLNQOVLDLKQKLASKA-----SAGADLKDIQSLRL 478
QY 289 ASQDWGKIPDESQRLLHTVLKPKWAKLCOVLHLSDACTSMVGNFLEYVIENHRIYEEPAT 348
Db 479 ENK---SLEDQQR-----VQAEKEVLQOQLSQTKARLEKVTETLTKNTASQKRDLETORN 530
QY 349 TFOAFQIALSWIAALLVKQI-LSHKESL-----VRANSELAFKCSRVEV-----391
Db 531 EWSKAKKVLLEEISRLKKEVDLARKNSAAPTIVEIKPEPNIAESITRVEQLMKEQKPOLP 590
QY 392 ----DVIYISILSCMKSL-----FLEHTQGLQFDCFGTNSKQSVVSTKLVNESLSGATVRDE 443
Db 591 LTARDFVQVLVAKLAKDAQGHKDLDASEDVIRECIG-----LLNRVLK-BPVFDE 638
QY 444 KIN-----TKSMRNSSEDEP--CWTEKRCSHYSTATRDIEKTISSGIKKYKQVQ 491
Db 639 KDKEMLQTLQNELNEALRKGKEDREQSAQLQSIIDKYEGEKESLEKEV-----KVHKEIK 694
QY 492 KLVQHEEKQKVELLN-----MYADKKQKLETSKSVAAVIRITCSRTSTOV---G 538
Db 695 KUKQBITERKTSYLANQKQERDMLKSEYEQIKNLNADHKKAAAEKVKHQNELTQLRKDG 754
QY 539 DLKLLDHNHYERKFDIEKSEKNECLKSLEBQMHDVAKKLAED 579
Db 755 DLK--ETNLQKLDLTLRQCNESERNRLQADYAAEKAKLTKD 793
```

Search completed: July 12, 2005, 01:07:32
Job time : 60.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 09:40:15 ; Search time 7459.33 Seconds
(without alignments)
4033.965 Million cell updates/sec

Title: SEQ3 D1219_COPY_650_1270

Perfect score: 3217
Sequence: 1 RADAFILGSSLNPSHDVXH.....SGNNKHFGSSNISONAPDV 621

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2_1/USPTO spool/ROBINSON929/runat_11072005_102152_29171/app query.fasta_1.2325
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=ROBINSON929 @CGN 1 1.12596 @runat_11072005_102152_29171 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3213	99.9	6554	8 AF213627	Arabidopsis
2	3213	99.9	6571	6 AX063583	Sequence
3	2977	92.5	10329	6 AX063582	Sequence
4	2975	92.5	10329	8 AF213628	Arabidopsis

C	5	2975	92.5	120965	8	AC026875	AC026875 Genomic s
	6	1033	32.1	695	8	ATH552638	Arabidops
	7	521.5	16.2	4027	8	AK068790	Oryza sat
	8	494.5	15.4	467	6	AX063609	Sequence
	9	469	14.6	515	6	AX063611	Sequence
	10	442	13.7	519	6	AX063607	Sequence
	11	397	12.3	184118	8	AP005621	Oryza sat
	12	332.5	10.3	150597	2	AP005005	Oryza sat
	13	332.5	10.3	156603	8	AP004752	Oryza sat
	14	164.5	5.1	4190	10	BC046638	Mus muscu
	15	164.5	5.1	7318	10	AB026497	Mus muscu
	16	164	5.1	5837	6	AR447512	Sequence
	17	164	5.1	7764	6	CQ414477	Sequence
	18	164	5.1	7764	6	AX828387	Sequence
	19	164	5.1	7764	6	AF006514	Homo sapi
	20	160.5	5.0	4769	6	CQ843036	Sequence
	21	160.5	5.0	4769	9	AK125995	Homo sapi
	22	160	5.0	3429	10	AB029919	Mus muscu
	23	158	4.9	7396	6	AX686226	Sequence
	24	158	4.9	110000	2	PFMAL8P1_11	Continuation (12 o
	25	157	4.9	4597	10	RNO237670	Continuation (12 o
	26	157	4.9	5946	6	CQ730200	Sequence
	27	157	4.9	5998	9	CR456526	Homo sapi
	28	157	4.9	7450	9	AB191263	Homo sapi
	29	156	4.8	5061	10	BC076567	Mus muscu
	30	156	4.8	5175	6	BD079934	Cancer-as
	31	156	4.8	5175	9	HUMCLIP	Human cytop
	32	156	4.8	6730	5	AY217130	Taeniopyg
	33	156	4.8	6945	5	AY217129	Taeniopyg
	34	155.5	4.8	95679	8	ATT29A15	Arabidops
	35	155.5	4.8	200001	8	ATCHRIV67	Arabidops
	36	155	4.8	110000	8	CR382139_15	Continuation (16 o
	37	154.5	4.8	38570	1	AC000108	Helicobac
	38	154	4.8	5176	6	CQ207060	Sequence
	39	154	4.8	5857	6	AR274942	Sequence
	40	154	4.8	5857	6	AX329796	Sequence
	41	154	4.8	5857	9	HSRESTIN	H.sapiens m
	42	154	4.8	5858	6	A21860	H.sapiens M
	43	153.5	4.8	6965	9	HSU31906	Homo sapien
	44	153.5	4.8	7636	6	AX933989	Sequence
	45	153.5	4.8	7695	6	AX336423	Sequence

ALIGNMENTS

RESULT 1	AF213627	Arabidopsis thaliana MOM mRNA, complete cds.	6554 bp	linear	PLN 01-JUN-2000
LOCUS	AF213627	Arabidopsis thaliana MOM mRNA, complete cds.			
DEFINITION	AF213627	Arabidopsis thaliana MOM mRNA, complete cds.			
ACCESSION	AF213627	Arabidopsis thaliana MOM mRNA, complete cds.			
VERSION	AF213627.1	GI:8132767			
KEYWORDS		Arabidopsis thaliana (thale cress)			
SOURCE		Arabidopsis thaliana			
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.			
REFERENCE		1 (bases 1 to 6554)			
AUTHORS		Amedeo P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.			
TITLE		Disruption of the plant gene MOM releases transcriptional silencing of methylated genes			
JOURNAL		Nature 405 (6783), 203-206 (2000)			
MEDLINE		20279299			
PUBMED		10821279			
REFERENCE		2 (bases 1 to 6554)			
AUTHORS		Amedeo P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J.			
TITLE		Direct Submission			
JOURNAL		Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland			
FEATURES		Location/Qualifiers			
source		1. .6554			
		/organism="Arabidopsis thaliana"			

APP 1

/mol_type="mRNA"		2437		GCTGTAGAGAACCTAAACCGCTCTCTCAGCGCAGCAGCTGCTCATGTGGGGGGCGTCATAC	2496
/cultivar="Zuerich"		81		LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
/db_xref="taxon:3702"		2497		TTATTTGATAACTGGATCAATTTTCACAGCAGTGAATCCAGATTCAGAGTTTCATTT	2556
/chromosome="1"		101		GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIleGly	120
/map="between ATEAT1 and phyA"		2557		GAACAATCTATTATGACGGCGTGATTCATGAATTCCTCGTCATATCTTCTTCCAAAGT	2616
310..6315		121		GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr	140
/function="involved in the maintenance of transcriptional		2617		GGAGAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGAACT	2676
gene silencing"		141		TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLeuLeuSerAspGluGluSer	160
/codon_start=1		2677		TACAGCAGTGATCTTACTCTATTTCGTGAAGCAATATTAAGTTCTCAGATGAAGAGT	2736
/product="MOM"		161		ProAsnIlePheTyrSerLysLeuGlyGlyLysAsnProMetTyrLysTyrProSer	180
/protein_id="AAF73380.1"		2737		CCAAATATATTTTGGTCAAGCTGTGGGGGAAAAAATCCCTATGTGAAATACCTCTCA	2796
/translation="MKDKKIGLGTCTIYTRSLAASIPASVQETPGLRRSRGTPST		181		AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
KVTTPASATRKSRLESPASVSKSGIVKNSPTSSLRNRGRKTEVLSQSKSGSDN		2797		GATACCTCCCAAGGAATCGAAAAACGAGTTCAGTATTTTGAGGGTTCGAAAGCGAGTCCC	2856
VVRGDSPIIDKSDVEESTDKPIIMASARYALPRGLKESALVDASPEEEL		201		LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAsp	220
LVKGSRRIPAGNDVQGTDCPPADAGSKRLPVDNLSKGTDFPLKSVTETKIV		2857		AAAATCTGGCGATGGTGGAAATGCAAAAGCGAAGAGAGGCTTCGTGATGTCTCAT	2916
LDASPIVETGDSVSGSPSENLETQKDGKDCPPANAESKTLPVGETSLEKYPQ		221		ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis	240
KFDQNDTCLPPANESKRLPVGETSLEKDTDFPLKSTTETGKWLIVASPIVETKDS		2917		CCCCGGTCACTGATCCGCCAGTAGATGATGATGAAGAAAGCCCTCTGGGAAGGATCAC	2976
VTPSGCKNDNRPMNCTCKRKVNGDCNRSVCITVQVPEBSDNVTQDMKETGP		241		MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly	260
VTSRYEENGQIOHGKSDPKFYSSVYPRYVPLVDSVQLBOYQCTLPFKSLSLSSL		2977		ATGGGGGGCTTTGGAGTCCACCAAGTCAATACACCTCCAGTCAATGATGATAATCTCTGT	3036
SKIDGALERTINVRKTCDDHPYVMDASLKQLTKNLELHEILDVEIKASGKLHLDK		261		ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle	280
MLTHIKNGKLKAVFYQATQTPGELLGNILEDFVQGRGPKSYEHEIYSSKKNAIN		3037		ACAGATGGTACATCGATGGAAATCATGCTTTTGGCTTGATTTCTATGGCAGCCATATC	3096
NFKNSQCCVLLLETRACQTKLIRADAFILFGSSLPNSHDVHKVEIKIESCSERT		281		SerGlyIleProGluAspMetLeuAlaSerGlnAspTyrGlyIleProAspGluSer	300
KIPRYSVCTVEBEKALILARQNMRQKAVENRSLTHALLMWSYLFDKLDHFFHS		3097		TCTGNAATCCCAAGAGATATGTTAGCTAGTCAAGATTTGGGGGAAAAATACCGGATGAATCA	3156
ETPDSGVSPEQIMDGIHEPSSILSSKGBENEVKLCILLEAKHAQGTYSDDSLFG		301		GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis	320
BOHILKSDDESNI FWSKLLGGKPMWKYPSTPOKRNKRVOYFEGSASPKTGDGNG		3157		CAGAGGAGGCTCCACACTGTTTTAAAGCCGAAGATGGCAAAACTTTGCCAGTTTTGCAT	3216
AKRKASDDVTDPRVTPDPPVDDDERKASGDKMGALSPKVTILOSCKSGTGTLL		321		LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis	340
DGNDAFGLYSMSHISIGIPEDMLASQDWGKI PDESQRRLHTVLKPKMAKLCQVHLSD		3217		CTTTCAGATGCTTGCAAGCATGCTCGGAAATTTCTCGAATATGTTATTGAAATACAC	3276
ACTSMVGNLEIRIYEPATTFQAFIALSWIAALVLKQILSHKESLVRANSE		341		ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle	360
LAFKSPSEVDYIYILSCMKSLFLEHTQGLQDFGNTSKQSVVSTKLVNESLSGAT		3277		CGAATCTCAGAAGAGCCAGCCACTACTTTTCAGGCAATTCAGATAGCCCTCGATGGATT	3336
VRDEKIKTSMNSDEDECKTCKSHYSTATRTDIETKISGIKKYKKYQKVLQVEH		361		AlaAlaLeuLeuValLysGlnIleLeuSerHisLysLysGluSerLeuValArgAlaAsnSer	380
EBKKWELLNMYADKKQLTSKSEVAARITCSRTSTQVGLDKLLDHNRYKRFBIK		3337		CGAGCTTGTGTGGTAAGCAAAATCTTAGCCCAAAAGAAATCTCTGGTCCGTCGCAATCT	3396
SEKNCLSKLEQHVAVKKLAEDACWLNRIKSWAALKVCVPIQSGNNKHFSGSN		381		GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys	400
ISQNRDVQVICNNANVEATYDTCNMAKVNQVPAENTLGTWSGGSGTQVHEMDVR		3397		GAATTAGCTTTCAATAGCTCTAGAGTAGAGTGGATATATTTATTCGATATTTGCTCTGC	3456
NDETMDVSALSRQUTQSNEHASITVPEILI PADCEFAALNVLHUSEDQNCORIT		401		MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn	420
SAASDEDVSSRVPESVLENSASPEFSNREELVTTENRTSRHVGFDPTDNIILDQ		3457		ATGAAGAGTCTGTTCTCGGAGCATACAAAGGTTTTGCAGTTTCGATTTGCTTCTTAAT	3516
NREDCSLDOEIPDELAMPVQHLASVVEVTRGAASDQYQDIPCMPSSLAGKQDPAAN		421		SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal	440
TSENLLEAI EPQASGETVVDIETMAHQDQVCTPLSSPTGNQAPAEANI EQONTV		3517		TCTAAACAGTCTAGTGGTTAGCACAAAACCTAGTAAATGAAGTCTCTCAGGGGCTACAGTG	3576
LNYTAPQLOQASFPAPALVAPLQLOQSPAPGAPAPLQOASFSFSSVSRPSALL					
LNFAVCPMPQRPQLISNTAPTSVTPATNPLRSPAPHLNSRPSSTTPVATATPTS					
SVYPPQALITYSAVISIQOQEQPQSSLSLSSGQSNEVVCLSDDDE"					
ORIGIN					
Alignment Scores:					
Pred. No.:	1,45e-249	Length:	6554		
Score:	3213.00	Matches:	620		
Percent Similarity:	100.00%	Conservative:	1		
Best Local Similarity:	99.84%	Mismatches:	0		
Query Match:	99.88%	Indels:	0		
DB:	8	Gaps:	0		
SEQ3_D1219_COPY_650_1270 (1-621) x AP213627 (1-6554)					
QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis	20		
DB	2257	CGAGCTGATGCGGTTATCTTTTGAAGCAGCTTGAATCCATCCGATGATGTTAAGCAC	2316		
QY	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40		
DB	2317	GTAGAGAAGATAAAAATCAGATCATGTTCTTGAAAGAACTAAGATATTCGATTGACTCA	2376		
QY	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys	60		
DB	2377	GTATGTACAGTTTGAAGAAAAGCCCTGATTTCTGGCTAGGCAAAATATCGGCAAAATAA	2436		
QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80		

```

QY 441 ArgAspGluLysLeuThrLysSerMetArgAsnSerSerGluAspGluGluCysMet 460
DB 3577 CGTGACGAAAGATTAAATACCAAGTCGATCGCAAGATAGCTCAGAGCATGAGAGTGCATG 3636
QY 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
DB 3637 ACTGGAAGAGATGTTAGCCATTATAGCACAGCAACAGAGATATCAAAAAGACTATTAGT 3696
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 3697 GGCATAAAAGAAATACAGAAGCAAGTCGCAAAAGCTTGTACAGAGCATGAGGAAG 3756
QY 501 LysMetGluLeuLeuLeuMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer 520
DB 3757 AAAATGAGCTGTAAATATATATGATGACAGCAAGCAAGCAAGCAAGCAAGCAAGT 3816
QY 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGluAspLeu 540
DB 3817 GTGGAAGCAGCAGTAATTCGTATTACTGTTTCACGACCAAGTACTCAAGTGGGTGATCTC 3876
QY 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluLysSerGluLysAsnGlu 560
DB 3877 AAATCTGCTGATCATTAATATGAAAGAAATTTGATGAATCAAAAGTGAAGAAATGAA 3936
QY 561 CysLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGluAspGlu 580
DB 3937 TGCCTCAAAAGCTCGAGCAAAATGCACAGGTTGCAAGAAAGATTTGGCTGAGGATGAA 3996
QY 581 AlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValProIle 600
DB 3997 GCCTGTGTGATTAATCGGATAAAGAGCTGGCAGCTAAATATAAAGTTTGTGTCCCAT 4056
QY 601 GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp 620
DB 4057 CAAAGTGGCAATACAGCATTTTAGTGTTCATCAACATTTCCCAAAATGCTCTGAT 4116
QY 621 Val 621
DB 4117 GTA 4119

RESULT 2
AX063583
LOCUS
DEFINITION
ACCESSION AX063583
VERSION AX063583.1 GI:12541307
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1
AUTHORS Habu, Y., Mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 010801-A 2 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source
location/Qualifiers
1..6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310..6315
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
/translation="MKDKIKIGTGTGTIVTSRLAASIPASVEQETPLRRSRGTPST
KVITPASATRKSERLAPSPASVSKSGGIVKNSPTSLRRSRNKGTEVLSQSKSGSDN
SRKGGTSPDIQRKDSVEEDTKIKPIMSARSYRALFRGLKESALVDAPNEEL
VVGCSRRIPAGNDVQGGTDCPPPADAGSKLPVDETSLDKGTDPLPKSVTETEKIV

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,46e-249 Length: 6571
Score: 2123.00 Matches: 620
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.84% Mismatches: 0
Query Match: 99.84% Indels: 0
DB: 6 Gaps: 0

SEQ3_D1219_COPY_650_1270 (1-621) x AX063583 (1-6571)

QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
DB 2257 CGAGCTGATGCGTTTATTCTTTTGAAGCAGCTTGAATCCCATCGCATGATGTTAAGCAC 2316
QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
DB 2317 GTAGAGAAGATAAAATCGAGTCATGTTCTGAAGAACTAAGATATTCGATTTGTACTCA 2376
QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
DB 2377 GTATGTACAGTTGAAGAAAACCCCTGATCTGGCTAGGCAAAATATGCGCAAAATATAA 2436
QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
DB 2437 GCTGTAGAGAACCTAAACCGCTCTCTCACGACGACTGCTCATGTGGGGGGCGTCATAC 2496
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
DB 2497 TTATTTTGAATAAATCGGATCATTTTTCACAGCAGTGAACCTCCAGATTCAGAGGTTTCATTT 2556
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly 120
DB 2557 GAACAAATCTATTATGACGCGCGTGATTCAATGAATCTCTGTCATATCTTTCTTCCAAGGT 2616
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
DB 2617 GGAGAAGAAATGAAGTCAAGCTGTGCTTACTTTTGGAGCCCAAGCATGCTCAGGGAAC 2676
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160

```

Db 2677 TACAGCAGTGATTCTACTCTATTGTTGGTGAAGCACCATTATTAAGTTGTGCAGATGAAGAGCT 2736
QY 161 ProAsnIlePheTyrSerLysLeuLeuGlyGlyLysAsnProMetTyrLysTyrProSer 180
Db 2737 CCANATATATTTTGGTCAAGCTGTTGGGGGNAANAATCCCTATGFGGAATACCCCTTCA 2796
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 2797 GATACTCCCAAGGAATCGAAGACGAGTTTCAGTATTTTGGGGTTCTGAAGCGAGTCCC 2856
QY 201 LysThrGlyAspGlyClyAsnAlaLysLysArgLysLysAlaSerAspValThrAsp 220
Db 2857 AAAACTGGCGATGGTGGAAATGCAAGAAGCGAAGAGCGCTTCGTATGATGTGCACTGAT 2916
QY 221 ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 2917 CCCCGGTCACGTATCCCGCAGTAGATGATGATGAAGAAGAGCCCTCTGGGAGGATCAC 2976
QY 241 MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly 260
Db 2977 ATGGGGCTTTGGAGTCACCAAAAGTCATAACACTCCAGTCCATCATGTAATCTTCTGCT 3036
QY 261 ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle 280
Db 3037 ACAGATGGTACATTTGGATGGAAATGATGCTTTTGGCTTGTATTTATGGGCGAGCCATATC 3096
QY 281 SerGlyIleProGluAspMetLeuAlaSerGlnAspTyrGlyLysIleProAspGluSer 300
Db 3097 TCTGGATCCAGAGATATGTTAGCTAGTCAAGATGGGGNAATACCGGATGATCA 3156
QY 301 GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis 320
Db 3157 CAGAGGAGCTCCACACTGTTTAAAGCCGAAGATGCAAAACTTTTGCCCAAGTTTGGCAT 3216
QY 321 LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis 340
Db 3217 CTTTCAGATGCTTGGCAAGCATGGTCCGAAATTTTCTCGAATATGTTTAAATGCAATCAC 3276
QY 341 ArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTyrIle 360
Db 3277 CGAATCTACAGAGAGCAGCCACTACTTTTCAGGCATTCAGATAGCCCTCGATGGATT 3336
QY 361 AlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer 380
Db 3337 CGAGCCTTGTGTAAGCAAAATCTTAGCCACAAAGAAATCTGTGTCGCGTCAAAATCT 3396
QY 381 GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys 400
Db 3397 GAATTAGCTTCAAAATGCTTAGAGTAGAGGGATTAATTTATTCGATATTTGCTCCTGC 3456
QY 401 MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn 420
Db 3457 ATGAGAGCTCTGTTCTCGAGCATAACAGAGTTTCAGTTCGATGCTTTGGTACTAAT 3516
QY 421 SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal 440
Db 3517 TCTAAACAGTCAGTGGTTAGCACAAACTAGTAAATGAAAGTCTCTCAGGGGCTACAGTG 3576
QY 441 ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerCysLysGluGluCysMet 460
Db 3577 CGTGACGAAGAAGATTAAATACGAGTGCATGCGGAAATAGCTCAGAGGATGAAGAGTGCATG 3636
QY 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 3637 ACTGAGAGAGATGTAGCCATTATAGCACAGCAACAGAGATATCGAAAGACTATTACT 3696
QY 481 GlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 3697 GGCATAAAAAAGAAATACAGAAGCAAGTGCAAAAGCTTGTACAGAGCATGAGGAAAAG 3756
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLysGlnLysLeuGluThrSerLysSer 520
Db 3757 AAAATGGAGCTGTTAAATATGATGTCAGACCAAGACGAAACTTGAATAAGT 3816

QY 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540
Db 3817 GTGGAGAGCAGCAGTAGTAATTCGTATTACCTGTTTCAAGGACAGTACTCAAGTGGTGATCTC 3876
QY 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
Db 3877 AAATCTGCTGGATCATAAATTTATCAAGAGAAAGTTTGTATGAATCAAAAGTGAGAAAAATGAA 3936
QY 561 CysLeuLysSerLeuGluGlnMethIleAspValAlaLysLysLysLysLysLysLys 580
Db 3937 TGCCTCAAAAGTCTCGAGCAAAATGCAGGAGTTGCAAGAGAAAGTTGCTGAGGATGAA 3996
QY 581 AlaCysTyrIleAsnArgIleLysSerTyrAlaAlaLysLysLysValCysValProIle 600
Db 3997 GCCTGTGGATTAATCGGATAAGAGCTGGGAGCTAAATTAAGTTTGTGTCCATT 4056
QY 601 GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp 620
Db 4057 CAAAGTGGCAATAAACAAGCATTTTAGTGGTTTCATCAAAACATTTCCCAAAATGCTCCTGAT 4116
QY 621 Val 621
Db 4117 GTA 4119
RESULT 3
AX063582 10329 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 1 from Patent WO0100801.
DEFINITION AX063582
ACCESSION AX063582
VERSION AX063582.1 GI:12541306
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
REFERENCE
AUTHORS Habu, Y., Mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 1 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
Location/Qualifiers
source 1. 10329
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
intron 1009..1295
intron 2551..2673
intron 2753..2867
intron 3114..3506
intron 3681..3973
intron 4896..4975
intron 5218..5777
intron 5883..6082
intron 7481..7615
intron 7772..7914
intron 8071..8153
intron 8319..8451
intron 8630..8718
intron 8919..9000
intron 9212..9284
ORIGIN
Alignment Scores:
Pred. No.: 3,38e-230 Length: 10329
Score: 2977.00 Matches: 617
Percent Similarity: 68.51% Conservative: 1
Best Local Similarity: 68.40% Mismatches: 3
Query Match: 92.54% Indels: 284
DB: 6 Gaps: 3

SEQ3_D1219_COPY_650_1270 (1-621) x AX063582 (1-10329)

QY	1	ArgAlaAspAlaPheLeuPheGlySerSerLeuAsnProSerHisAspValLysHis	20	322	-----	322
Db	4172	CGAGCTGATGCGTTTATCTTTTGGAGCAGCTTGAATCCATCCGATGATGTTAAGCAC	4231	5250	TTAGCCTTCATATGCTTATATAGCTGATTGCAACTGTAGTTGTGTACCTGATTTCCCTG	5309
QY	21	valGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40	322	-----	322
Db	4232	GTAGAAGAGATAAAATCGAGTCATGTTCTGAAGAAGAACTAAGATATTTCCGATTGTACTCA	4291	5310	TTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTA	5369
QY	41	ValCysThrValGluGluLysAlaLeuLeuLeuAlaArgGlnAsnMetArgGlnAsnLys	60	322	-----	322
Db	4292	GTATGTACAGTTGAGAAAAGCCCTGATTCTCGCTAGGCAAAATATGCGCAAAATATAA	4351	5370	TTTCTTATATCTGGCTTCGTTACCAATGCACCTGTTAAATGAGCAACTGCTGCACAAAC	5429
QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80	322	-----	322
Db	4352	GCTGTAGAGAACCTAAACCGCTCTCTCACGACGCACTGCTCATGTGGGGGGCTCATAC	4411	5430	AGTAGGTATATGTGCCTCATGTCATTCATTTGTTTATTAAGCAAAAGAAATTTCTGTCT	5489
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100	322	-----	322
Db	4412	TTATTTGATAAATCGATCATTTTTCACAGCAGTGAACCTCCAGATTCAGAGGTTTCATTT	4471	5490	ACTTTACATGATCCATCTCTGGGAGTATATACTATATAAACCCTTAGGCCCTTTGTACCT	5549
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly	120	322	-----	322
Db	4472	GACAAATCTATTATGACGCGGTGATTCAGAAATCTCTGFCCTACTTTCTTCCAAAGGT	4531	5550	GGCTGATCAAGACATGTCAAAGTTTATCTGTTCCCTGTTGGTATAGAACTAATACAG	5609
QY	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140	322	-----	322
Db	4532	GGAGAAGAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGNACT	4591	5610	TGCTGATGCTATTTTAAAGGTAGTCTTATGTCTTACATATTTGGCTAATAGATGTTTCCG	5669
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160	322	-----	322
Db	4592	TACAGCAGTGATTTCTACTCTATTGTTGTAAGACCATATTAAGTTGTTCAGATGAAGAGT	4651	5670	CTGCTGTCATATATCTTCTGTGATTTATCACGGTCTCCGCTCTATCAAAATTTGTACTAA	5729
QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpProSer	180	323	-----	326
Db	4652	CCAAATATATTTTGGTCAAGCTGTTGGGGGAAAATACTTAATGTGGAAATACCTTTCA	4711	5730	AGGTATTTTGGCAATGTGTGATTGGTTAACAGATTTATTTGTTTTCAGATGCTTCACAC	5789
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro	346
Db	4712	GATACTCCCAAGGAATCGAAACGATTCAGTATTTTGAGGGTTCTGAGCGAGTCCC	4771	5790	AGCATGTCGGAATTTTCTCGAATATGTTATTTGAAATCACCAGATCTACGAAGAGCCA	5849
QY	201	LysThrGlyAspGlyLysAlaLysLysArgLysLysAlaSerAspValThrAsp	220	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu	357
Db	4772	AAAACCTGGCGATGCTGGAATGCAAGAAAGCGAAGAGCTTCTCATGATGTCTACTGAT	4831	5850	GCCACTACTTTTTCAGCATTCAGATGAGCCCT - GGTAACAGCAGCATTTACTTTGATAAT	5908
QY	221	ProArgValThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHis	240	357	-----	357
Db	4832	CCCCGGTCACTGATCCGCCAGTAGATGATGATGAAGAAAGGCCCTCTGGGAAGGATCAC	4891	5909	TATGCAATGTTTCTTCATCATCTGCTTTGTTTGTAGAAATGTCCTCAGAAGGCAGACTCC	5968
QY	241	Met-----	241	357	-----	357
Db	4892	ATGGGTAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA	4951	5969	TTTAGTTTAACTTTTCCAATCATAGGATTCAAATATCCATTAATGCTGCTTTCGATCGCTG	6028
QY	242	-----	253	358	-----	359
Db	4952	CTTTGCTGACTATCTTCTTACAGGGGCT - TTGGAGTCAACAAAGTCATAACACTCCAG	5010	6029	CATAATATATGAATAGTTGACATACACTGTAATACCTGTTGTTTAAATATGCAATTTTCAGAGTGG	6088
QY	254	SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273	360	IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn	379
Db	5011	TCATCATGTAATTCCTTCTGTTACAGATGTTACATTTTGGATGGAATATCTCTGCTCGTGC	5070	6089	ATTGACGCTTGTGTTGTAAGCAATTTCTTAGCCCAAGAAATCTCTGCTCGTGCATAT	6148
QY	274	TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293	380	SerGluLeuAlaPheLysCysSerArgValGluValAlaAspTyrIleTyrSerIleLeuSer	399
Db	5071	TATTTCTATGGGACCATATCTCTGGAAATCCAGAGGATATGTTAGCTAGTCAAGATTGG	5130	6149	TCCTGAATTAGCTTCAAAATGCTCTAGAGTAGAGTGGAATTATATTTATTCGATATTGTC	6208
QY	294	GlyLysIleProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetAla	313	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419
Db	5131	GGGAAAATACCGATGATGATACACAGAGAGGCTCCACACTGTTTAAAGCCGAAGATGCA	5190	6209	TGCTGAAGAGTCTGTCTCTGGAGCATACACAGGTTTTCAGTTTCGATTGCTTGTACT	6268
QY	314	LysLeuCysGlnValLeuHisLeuSer-----	322	420	AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr	439
Db	5191	AAACTTTGCCAAGTTTTCATCTTTC - AGTAAGTGGCCTTTTTCACCTCCACAACTTATT	5249	6269	AACTTAAACAGTCACTGTTGTAGCACAACAACTAGTAATAAGAAAGTCTCTCAGGGGTACA	6328

QY	440	ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys	459
Db	6329	GTGCGTGACGAAAGATTAAACGAAGTCGATCGCAATAGCTCAGAGGATGAAGAGTGC	6388
QY	460	MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle	479
Db	6389	ATGACTTGAGAGAGATGTAGCCATTATACACAGCAACAGAGATATCGAAAGACTATT	6448
QY	480	SerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu	499
Db	6449	AGTGGCATAAAAAAGAAATACAGAAGCAAGTGCAAAACCTTGTAACAGGATGAGGA	6508
QY	500	LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys	519
Db	6509	AAGAANAATGAGCTGTAAATATGATATGACAGCAACAGCAAGCAAACTTGAATAGTAA	6568
QY	520	SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539
Db	6569	AGTGTGAAGCAGCAGTAATTCGTATTACTGTTACGACCAGTACTCAAGTGGGTGAT	6628
QY	540	LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn	559
Db	6629	CTCAAACTGCTGATCATATATGAAGAAAGTTTGATGAATCAAAAGTCAGAAAAAT	6688
QY	560	GluCysLeuLysSerLeuGlnMetHisAspValAlaLysLysLysLysLeuAlaGluAsp	579
Db	6689	GAATGCTCAAAAGTCGAGCAAAATGCAGAGTTGCAAGAGAAAGTGGCTGAGGAT	6748
QY	580	GluAlaCysTrrPileAsnArgIleLysSerTrrAlaAlaLysLeuLysValCysValPro	599
Db	6749	GAAGCTGTGTGATTAAATCGGATAAAGAGCTGGCAGCTAAATTAAGTTGTGTCCC	6808
QY	600	IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro	619
Db	6809	ATTCAAGTGGCAATAACAAGCATTTTAGTGGTTCATCAACATTTCCCAAAATGCTCT	6868
QY	620	AspVal 621	
Db	6869	GATGTA 6874	

RESULT 4	AF213628	10329 bp	DNA	linear	PLN 01-JUN-2000
LOCUS	Arabidopsis thaliana MOM gene, complete cds.				
DEFINITION	AF213628				
ACCESSION	AF213628.1	GI:8132769			
VERSION					
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 10329)				
AUTHORS	Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.				
TITLE	Disruption of the plant gene MOM releases transcriptional silencing of methylated genes				
JOURNAL	Nature 405 (6783), 203-206 (2000)				
MEDLINE	20279299				
PUBMED	10821279				
REFERENCE	2 (bases 1 to 10329)				
AUTHORS	Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland				
FEATURES	Location/Qualifiers				
source	1..10329				
	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				
	/cultivar="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/map="between ATEAT1 and phyA"				

mrna					
	join(705..1008,1296..2550,2674..2752,2868..3113,3507..3680,3974..4895,4976..5217,5778..5882,6083..7480,7616..7771,7915..8070,8154..8318,8452..8629,8719..8918,9011..9211,9285..10047)				
	/product="MOM"				
	join(1301..2550,2674..2752,2868..3113,3507..3680,3974..4895,4976..5217,5778..5882,6083..7480,7616..7771,7915..8070,8154..8318,8452..8629,8719..8918,9001..9211,9285..9808)				
CDS	/function="involved in the maintenance of transcriptional gene silencing"				
	/codon_start=1				
	/product="MOM"				
	/protein_id="AAF73381.1"				
	/db_xref="GI:8132770"				
	/translation="MKDEKIGLTGRTIYTRSLAASIPASVEQETPGLRSSRGTPSTKIVTPASATRSKRLAPSPASVSKSGGIVKNSSTPSSLRNRRGKTEVSLQSKSGDN SIRGDTSPDIEQRKDSVEESTDKIPIMARSARYALFRGLKKESEALVDASPEEEL VVCGSRIPAGNDVQKTDICPPADAGSKRLPVDETSLDKDTDFPKSVTETEKIV LDASPIVETGDSVIGSPENLETKLQDGKDCSPANAESKTLVPGETSLEKEIPQ KFQDNTDCLPPANAESKRLPVGETSLEKDTDFLKSTTETGKMWLYASPIVETRDDS VICPSGTNLETKLLVSKTGLTDLVLPKRRDRTABTELDACATVANGDDHVMSSDG VIPSOGCKNDRPMECNTCKKROKNGDCNRSCVICOPVBESDNVTOOMKSTGP VTSREYENGQIQHGKSDPKFYSVYVPEYVQVLSVQLEOYCOTLIESKLSLSL SKIDIGALEETLNSVRKTCDRPYVNDASLKOLLYKNLSLHEILOVEIKASGKLHLDK MLTHIKKNGLRKAVVYQATQTPPEGLLLGNILEDFVQRFPGKSYEHGIYSSKNSAIN NFNKESQCILLLETRACQTKILLRADAFILFGSLNPSHDVHKHVEKIKIESCSERT KIFLYSVCTVEERAKILARQNKQKAVENLRSLTHALLMWGASYLFDKLHDFHSS ETPDGSVFEOISIMDVTHFESSILSSKGGENEVKLLLEAKHAQCTYKSDSTLFG EDHILKDEESPNIFMSKLGKGNPMKYPSDTQQRNKRKYVPEGSASPKYCTGGGN AKRKKASDDVDFRVDTPPVDDDRKASGDKHMGALSPKVIITLQSCSKSGTGTG DGNDAFLYSGMSHISGIPEDMLASQDMGKIPDESQRRLHTVLKPKMAKLCOVLHSD ACTMVGNGFLYIENHRIYEPATTFQAFQIALSWIAALLVKQLSHKESLVRANSE LAFKCSRVEDYIYISLCKMSLFLHQGLQFCFGTNSQSVSTKLNVLSLGSAT VRDEKINTKMRNSEDDECKTEKSHYSTARDIENTIGIKKKYKQVOKLQVEH EEKMWELLMYADKKOKLETSKSVEAAVIRITCSRTSTQVGDGLLDHNYERKDFEIK SEKNECLSLQEMHDVAKKLADEACWINRIKSWAKLKVCYVPIQSPNQAPANEI EQNI 1SQNAPDQVICNNANVEATYADTNCMAKSNQVPAENTLGTMSGTQQKHVMDVDR NDETMDSALSERQITKQSNESHASITVPEILIPADCEEFAALNVHLSDEQDCRIT SAADESDVSRVPEVSQLENLSASPEFSLNREALVTENRTSHVGFDTDNILDQO NREDCSLDOEIPDELAMPVHLSAVVETRGAAESDQYQDICTPMPSSLAGKQDPDAAN TESNLEAIPQSNAGSETVETDFAASHQGDQVTCPLLSPTGNQAPANEI EQNVTIV NTSAPHVAGDPDAGESDIYVIDQETMGQAQACSLPSGSGVGTQSLDGLANIEQNVTIV AQLPTGSDAVVTGSGVSDQCADSPMLSPGNHPTDVTNIGLNTQSNQVSVAEHVIS GSDACEMEISIEFGPQVERSTFANLFHEGGVSHSAGVTALVPSLNNGTQEQIAPVQO IPFPVFNDFLHELEKLRRESEKKTPEEKSKILKAELERMAVQAEFRKPFHEVE AEHNRTTKIEKDKNLVIMKLANAFSLCTDKKVPSPGAPRGKIQQLAQAQAQVSA LRNIAPQLOCASSFPAPALVAPLOQSSFPAPGAPLOPQASFPSSSPSSVRRPALL LNFACVMPQPROPLISNIAPTSPVTPATNPLRSPAPHLNSYRSPSSSTPTVATATPS SVPPQALYSAVSIQQQEQPQSSLSGLSNNEVCLSDDE"	4,9e-230	Length: 10329		
		2975.00	Matches: 617		
		Percent Similarity: 68.40%	Conservative: 0		
		Best Local Similarity: 68.40%	Mismatches: 4		
		Query Match: 92.48%	Indels: 284		
			Gaps: 3		
	SEQ3_D1219_COPY_650_1270 (1-621) x AF213628 (1-10329)				
QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisAaspValLysHis	20		
Db	4172	CGAGCTGATCGCTTTATTCTTTTGGAAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	4231		
QY	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40		
Db	4232	GTAGAGAAGATAAAAAATCGAGTCATGCTTCTGAAGAAGCTAAGATATTCGATGTGACTCA	4291		
QY	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys	60		
Db	4292	GTATGTACAGTTGAAGAAAAAGCCCTGATCTCGCTAGGCAAAATAAGCGCAAAATAAA	4351		

QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80	QY	322	-----	322
Db	4352	GCTGTAGAGAACCTAAACCGCTCTCTCAGCAGCAGCTGCTCATGTGGGGGGCTCATAC	4411	Db	5430	AGTAGGTAGTTATGTGCTCCTCATGTCTATTCATTGTTTATTGAAGCAAGAAATTTCTGTCT	5489
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100	QY	322	-----	322
Db	4412	TTATTTTGATAAACTGGATCATTTTTCACAGCAGTGAACCTCCAGATTGAGAGTTTCATTT	4471	Db	5490	ACTTTACATGATCCATCTGTGGAGTATATACTATATATAAACCCTTAGGCCCTTTGTACCT	5549
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120	QY	322	-----	322
Db	4472	GAACATCTATTATGACGGGTGATTTCATGATTCTGTCATATCTTCTTCCAAAGGT	4531	Db	5550	GGCTGATCAAGACATGTCAAAGTTTATCTGTGCTGTTGGTATAGAAACTAATACAG	5609
QY	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140	QY	322	-----	322
Db	4532	GGAGAAAGAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAACT	4591	Db	5610	TGCTGATGCTATTTTAAGGTAGTCTTATGTCTTCAATATTGGCTTAATAGATTTCGG	5669
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer	160	QY	322	-----	322
Db	4592	TACAGCAGTGATTCTACTCTATTGTGTGAAGACCATATTAAAGTTGTACATGAAGAGT	4651	Db	5670	CTGTGCTGCCAATATACCTCTGTGATTATCAGCGTGCTCCGTCTATCAAAATTTGTACTAA	5729
QY	161	ProAsnIlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSer	180	QY	323	-----	326
Db	4652	CCAAATATATTTGGTCAAAAGCTGTGGGGGAAAAAATCCTATGTGGAATACCTCTCA	4711	QY	5730	AAGGTATTTTGCATGTGTGATTGGTTAACAGATTATTTTGTTCAGGATGCTTCACACA	5789
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200	QY	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro	346
Db	4712	GATACTCCCAAGGAATCGMAACGAGTTTCAGTATTTTGAGGGTTCGAAGCGAGTCCC	4771	Db	5790	AGCATGGTCGAAATTTTCTCGAATATGTATTGAAATCACCAGATCTACGAAGAGCCA	5849
QY	201	LysThrGlyAspGlyGlyAsnAlaLysArgLysLeuAlaSerAspValThrAsp	220	QY	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu	357
Db	4772	AAAACTGGCGATGGTTCGAATGCAAGAACGCAAGAGAGCTTCTCATGTATGCTACTGAT	4831	Db	5850	GCACACTACTTTTCAGGCATTCAGATAGCCCT- GGATGACAGCATTTTACTTTGTGATAAT	5908
QY	221	ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis	240	QY	357	-----	357
Db	4832	CCCCGGTCACTGATCCGCCAGTAGATGATGAAGAAAGAGCCCTCTGGGAAGGATCAC	4891	Db	5909	TATGATTTGTTTCTTCATCATCTGCTTTTGTAGTAGTCTCCTCAGAGCGACACTCC	5968
QY	241	Met	241	QY	357	-----	357
Db	4892	ATGGGTAAATAAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA	4951	Db	5969	TTTAGTTTTTAACCTTTTCCAATCATAGGATTCAAATATCCATTAATCGCCCTTTGATCGCTG	6028
QY	242	-----	253	QY	358	-----	359
Db	4952	CTTTGCTGACTATCTTTCTACAGGGCT- TTGGAGTCAACCAAAAGCTATAACACTCCAG	5010	Db	6029	CATAATATATGAATAGTTGACATCTAGTAACTAGCTGTTGTTAATATGCAATTTTCAGAGTTGG	6088
QY	254	SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273	QY	360	IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn	379
Db	5011	TCATCATGTAAATCTTCTGGTACAGATGGTACATTCGATGGAATGATGCTTTTGGCTTG	5070	Db	6089	ATTGAGCCCTCTTGGTAAAGCAATCTTAGCCACAAAGATCTCTGTGCTCGTCAAAAT	6148
QY	274	TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293	QY	380	SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer	399
Db	5071	TATTTCTATGGCAGCCATATCTCTGGAATCCACAGAGATATGTTAGCTAGTCAAGATTGG	5130	Db	6149	TCTGAATTAGCTTTCAAATGCTCTAGAGTAGAGTGGATTATATTTATTCGATATTGTCC	6208
QY	294	GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla	313	QY	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419
Db	5131	GGAAAAATACCGATGAATCACAGAGAGGCTCCACACTGTTTAAAGCCGAAGATGGCA	5190	Db	6209	TGCTGAAGAGTCTGTCTCTGGAGCATACACAGGTTTCAGTTCGATTGCTTTGTGACT	6268
QY	314	LysLeuCysGlnValLeuHisLeuSer	322	QY	420	AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr	439
Db	5191	AAACTTTGCCAAGTTTTCATCTTTC- AGTAAGTGGCCCTTTTCACCTCCACACTTATTT	5249	Db	6269	ANTTCTAAACAGTCAGTGGTAGCACAAACTAGTAAATGAAGTCTCTCAGGGGCTACA	6328
QY	322	-----	322	QY	440	ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys	459
Db	5250	TTAGCCCTTCATATGCTTATATATAGCTGATTCGAACCTGATGTTGTTTACCTGATTTCTCG	5309	Db	6329	GTGCGTGACAAAAGATTAAATACGAAGTCCGAAATAGCTCAGAGGATGAAGAGTGC	6388
QY	322	-----	322	QY	460	MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluYerThrIle	479
Db	5310	TTACAGCCAAATGTGAGAGTTTTTATTTCTTCAACTATATCCATCCGTTTAAGCATATTTTA	5369	Db	6389	ATGACTGAGAAAGAGATGTAGCCATTATAGCAGCAACAAAGAGATATCCAAAAGACTATT	6448
QY	322	-----	322	QY	480	SerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu	499
Db	5370	TTTCTTATATCTGGCTTGTTCCTAACATGCACTGTTTAAATGAGCAACTGCTGCACAAAAC	5429	Db	6449	AGTGCAATAAAAAGAAATACAAAGCAAGCAAGTGTGCAAAAGCTTTGTACAGAGCATGAGAA	6508
QY				QY	500	LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys	519

6509	Y	AGAAAAATGGAGCTCTTAAATATGTATGATCAGACCAAGACAGAAACCTTTGAACCTAGTAAA		6568
520	Y	SerValGluAlaAlaValIleargIleThrCySerArgThrSerThrGlnValGlyAsp		539
6569	Y	AGTGTGGAAAGCAGCAGTAAATTCGTATTAACCTGTTTACGACGACAGTACTCAAGTGGGTGAT		6628
540	Y	LeuLysLeuLeuAspHisAsnTyrrGluArgLysPheAspGluIleLysSerGluLysAsn		559
6629	Y	CTCAAACTGCTGGATCATATTAATGAAAGAAAGTTTGATGAAATCAAAAGTGAGAAAAAT		6688
560	Y	GluCysLeuLysSerLeuGluClnMethHisAspValIalLysLysLeuAlaGluAsp		579
6689	Y	GAATGCCCTCAAAAGTCTGGAGCAATGACCGATGTGTTCACAAAGAAAGAGTTGGCTTGAGAT		6748
580	Y	GluAlaCysTyrTolIleAsnArgIleLysSerTyrPalaalLysLeuLysValCysValpro		599
6749	Y	GAAGCCTGTGTGATTAAATCGGATAAAGAGCTGGGCAGCTAAATTAAGATTTGTGTTCCT		6808
600	Y	IleGlnSerGlyAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro		619
6809	Y	ATTCAAGTGGCAATAACAGCATTTTAAAGTTTCAATCAACATTTCCCAAAATGCTCCT		6868
620	Y	AspVal 621		
6869	Y	GATGTA 6874		
RESULT 5				
AC026875/c				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL		</		

Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Apr 27, 2000 this sequence version replaced gi:7543638.

Location/Qualifiers

1. 120965

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/db_xref="taxon:3702"

/chromosome="1"

/clone="T6D22"

join(410..559,596..758,828..865,1681..1731,1903..2253

3562..2999,3096..3346,3240..3322,3493..3659,3781..381

2992..4021,4069..4246,4597..4703,4824..4952,5058..512

5239..5394,5549..5709,5812..5843,6103..6266,6382..646

6658..6723,6881..7170,7235..7264,7463..7570,7663..777

7939..7999,8364..8440,8552..8623,8702..8778,8898..897

9337..9439,9593..9751,9883..10281)

/note="similar to AT.1.24-9 gene product gb|AA07881.1;

similar to EST emb|Z35021.1"

/codon_start=1

/evidence=not experimental

/product="T6D22.1"

/protein_id="AAF79821.1"

/db_xref="GI:8778822"

/translation="MIGHNFCGRNRLDSLFLPTTKNSLSQTI VAVPKISFVNRCS

TQTVKILQTGIDSGSPNQADKGGCKRVAAASRVCGGVLLRLINQKLNKPERD

LTYNKVIITQIGDGFMLVDEPDGFSFLASKPIDFCLLRWGSFVLLRSTNSVVPV

SQVSDSRRRPNLVAALSLVYVVRTKLSHVLSLI PHTTHPAFSINILHLSATS

SLRSEPTITLARLSLFCNSRIFLSSDMPKKQKQRHAEQKQVKKMPADPESGDS

SATVAAEVNNOFGSLKESNTNAPLPSITNHRVQNLVWPKPSYGTVGSSSAT

EVGTSAVSQIGSSGDTKVLNLSKIFGNLLEKFSVDKSTYCHAQIRFYPKFENE

KTDQIRIRMTLMEWSKGLATLEVLKSHGSLPMYAGHKGAVAKNSFGNIFIVSQAQ

VLLADLDVTFMDKIPHIITVVRQIYFKKHFDDEANARYYSFSLMLRLGEURLQK

KQHCQRDPLDYNSKAVQLKGNFDLALVVVAVTELGNCKPQVYSSEIISFRKQ

WRLPTNHWLFSRKSVSFFAAGFDALCEEATISVCRALDEVADISVPASKDHVKVQ

GEIILLVAVSVSSQSDRMENVLRDHPDQPCDGNLDGLSREI CAAHRSNEKQOM

RALLRSVGSPFCPSDFGVDESHPSKADSPYITKPLQPADYSTKLQEMVLMKE

KLPAAFCYKHNHRAEIDIPNLFPVQLVHVHSDSGFRYRHEKMHMSLWPLYRFL

FVDNLKSNKGRDLMALKSIDNASENDGSGDLADDANILMIKWKFLTYKLRFL

IRNGLSILPKOGAAAYTYTLRQMKGTSQKQELCKWLDWEAY IRKCKGNDOLS

STSYLSEAPLEQTAQSPKPHHILIGASNLVRKTEFLAI VDGDDDEGDLVKKQGS

TPATPEAPAKVAGQDEGLIFVFPYSYIELMDGSPGPSAKSLKELLNAPQGGD

RPVHTLMDGLVKGYTPKVAADERRKKPOSIMLAKNAPNEVDNRQITDCMRRTASAG

PIVADSGDTNTPYSLDALAMFVRQLRVNHGDKDESSNAGYLLMFPYHLYEGNQ

RNEPESLIERFSGLSIKMPLKSDRTPDPVKSVLEGI DLFLNLRHRRGLSETKR

TYAAETWKEQLRDTLVANSEYLSIIQVPPFSMHQVREELTKIAGDYKPPSSER

KHGSIVFAAINLPATVLSLEKLAANPFLSKGKKSIOEKRLERSHVLTAHKRS

HGVTAVKSYQHLNRENVPELTSLINDKMAALTARHVSVDGETTVVSKNEWHVTLWT

AEGVTAKEANTLPQLYLEGKASRLVIDPVSISGLEFF"

join(11454..11915,12009..12885,12925..13073,14200..142

14754..14777,14903..15370,15466..16353)

/note="putative elongation factor 1-a

/gb|AF13128.1|Af124801.1; similar to ESTs gb|AW398653.1,

emb|Z17990.1, gb|AW220334.1, gb|AW622036.1,

dbj|AV439770.1, gb|T45275.1, and gb|AW267999.1,

gb|T04414.1, gb|AW221780.1"

/codon_start=1

/evidence=not experimental

/product="T6D22.2"

/protein_id="AAF79822.1"

/db_xref="GI:8778823"

/translation="MGKEKFKHRIINVIIGHVDSGKSTTGHLYLGLGIDKRVIFERKE

EAEENKRSFKYAWLDKLKAERERGITIDIALWKFTTKYCTVIDAGHRDFIKM

ITGTSQACAVLI IDSTTGGPENGISKQQTREHALLAFLGWKIMCCNKMDATP

KISKARYDEIILKEVSYLUKYGNDPKPIFPVIFSGEGDNMIERNSTLNDWYKQPTLLE

ALDQINERPKSDKPLRLPLQDVYKIGGITVPGVGTGMIPKGMVTVFAPGTLTTE

VKSMHEHSLLEALPGDNVGNFNKVAANVKDLRQGVASNSKDDPAKGAANFTSQVIL

MNHQSIGNGVAPLVDCHTSHLAVFSELTIKIDRRSGKEIKEKPLKNGDAGMVKM

TPTKPMVETSEYPLGRFARVDRMROTVAVGIVKSDKKDPTGAKVTAAYVKVGSK

DSVEFITIWIWLCSCFVPSPERCSRNLVLRWRVLYTHLIPHHPSVCLDILVC

MEIHNFTVWGKEKFNHINVIIGHVDSGKSTTGHLYLGLGIDKRVIERFERKEA

MNKRSPKYAWLDDKLKAERERGITIDIALWKFTTKYCTVIDAGHRDFIKM

SQADCAVLI IDSTTGGPENGISKQQTREHALLAFLGWKIMCCNKMDATPKYS

ARYDELIKEVSSYLKVGNPKIPFVPIISGPEGNWIERSTNLDWYKPTLLEALDQ
INPKRSPDLPLQDVYKIGIGITVPVGVETGMKIPGMVVTFAPTGLITTEKSV
EMHESLALPQDFGVNFVKNVAVKDLKRGYVANSKODPAKGAANFTSQVIMNHP
GOINGVAPVLDCHTHIAVFKSEIITKIDRRSGKEIEKEPKLNGDAGWVKMTPTK
PMVETPSEYPLGRFAVRMROTVAVGVKISVDDKDPGAKVTKAAVKKGAK"
complement (join(19245, .20132, 20226, .20687))
/notes="putative elongation factor 1-a
gb|AAFI3128.1|AF124801.1; similar to ESTs gb|AW034714.1,
gb|AI812185.1, emb|217990.1, gb|AW267999.1, gb|N96009.1,
dbj|JAV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"

CDS

/translation="MGKEKFINIVIGHVDSKSTTGHLYIKLGGIDKRVIERPEK
EAAEMNRKPKYAWLDKLAEREGITIDIALWKETTKYCTVIDAPGRDPIKNM
ITGTSQADCAVLIIDSTTGGFEAGISKDGOTREHALLAFTLGWKMICCCNKWDATP
KYSKARYDRIIEKVSYLKVGNPKIPFVPIISGPEGNWIERSTNLDWYKPTLLE
ALDOINERKPSDKPURLPLQDVYKIGIGITVPVGVETGMKIPGMVVTFAPTGLTE
VKSVEHHESLLEALPQDFGVNFVKNVAVKDLKRGYVANSKODPAKGAANFTSQVIM
MHPGQINGVAPVLDCHTHIAVFKSEIITKIDRRSGKEIEKEPKLNGDAGWVKMT
PTPKPMVETPSEYPLGRFAVRMROTVAVGVKISVDDKDPGAKVTKAAVKKGAK"
complement (20787, .20857)
/product="trna-Lys"
/notes="codon recognized: AAA; putative trna-Lys"
/evidence=not experimental
complement (join(21956, .22189, 22535, .22600, 22713, .22929,
23041, .23126))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"

tRNA

/translation="MCLNRLPSATPIPRAEEREECSMGFRNHQLLLSSFLCKRPA
IIMNKGSGSGSGSGGPTAAALQKALQORVETDITSVVDNFTQIVNSRV
SPPVKNSETYMERASRMVQADSLKLVSELKQTAIFSGFASLNDHVEQRIEEF
DQEAETNRLARLADDSANLKELESHTYSSAQLTLDI"
join(23576, .23689, 24080, .24258, 24347, .24522, 24635, .24756)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"

CDS

/translation="WTGLARLVAPMILLIPIELVKAEVITLTPETPSDKLALSN
KAQLQNCISWCSFVSLSDTRCLLSPCLLLVQIKEDTAFWPCVPCWKRCKKJG
NLWEDLGKAMEGDEIEVGEVDCGTSRAVCTKVEIHSYPTFLMFLNGEVSRYKKGKD
VESLKAFFVVEETEKAERKAQLEDKEL"
complement (join(25423, .25563, 25675, .26048, 26127, .26330,
26585, .26783, 26941, .27842, 27936, .27991, 28208, .28413))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"

CDS

/translation="WDSVRRKINNTQSSPSPKPSKESVVRNPAALAASTANSIRP
SKSLVIFLLISAFSLISFEMAGEKWLITNALTFGKISQEAAYVTVKAWQGLVTFPC
IGAMWALSGILSHRKAFAARGETKSFKTDQFSLNRQLELLGKKKADQSVSEF
PKSRPLPAQSELPVTHQTLTGAHKSAGGDKLNSRNGSQISFSPKSPKSGE
PSMVLVPSSPVSNRLSSODKAVSPWSGRRSSAKDITTEOLOELLAEIDDKITE
SAGKWTPTPTVGSFAMASPTVGGSTGASGATRTPLRPVMSPCAQKFTPPKGE
GFPNPMISLERALGPHLGVTPQIEDWRDLRQWSSVLLKPLNKVETSHIQWQT
ASGLKGVNIVSGSLDPTNGTATTPALPDVTRKQWPSYSLDELDALLHQLRANLVQAI
DASMQKRTENQFOQQQQQQQQAALIPWQECVDATSEHRLQGLMKGEWVKGLP
RSSIPADYTVQIRALAEQTVKNYNGADTRKNKKWSLEPTDLSHLLYLFCAI
LEHPKMLHLDPSSVYGTQASKNPLFLGVLPPKRPPEKYIAVVSVPSTLHPGACVL
AVDKQSPPTFALYWDKKVQFTLQRTALWDSMLLICHRIVKVGYGGVVRGMILGSSALN
ILQVVDSDTDD"
complement (join(29418, .29491, 29579, .29620, 29858, .29942,
30386, .30805))

CDS

/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"

/translation="WVSSKKPKPKKARSVVVNKASGRSKRSSGSRTKTTSKNKIVK
KKPEIYIESSSSDSVEAIRGDEAKNSGVVSKNGSKSGVIGPITKSKRREDGG
AEDAKIKFPMNRIRRMISDNGAPQIMQDAVFLNKATMFIERFSEEAIVSSVYDKK
KETHYKHLSSVVSNDQRYEFLADSVPEKLKAEALAEWERGMTAG"

Alignment Scores:
Pred. No.: 1,4e-228 Length: 120965
Score: 2975.00 Matches: 617
Percent Similarity: 68.40% Conservative: 0
Best Local Similarity: 68.40% Mismatches: 4
Query Match: 92.48% Indels: 284
DB: 8 Gaps: 3

SEQ3_D1219_COPY_650_1270 (1-621) x AC026875 (1-120965)

QY	1	AtgAlaaspAlapheIleLeuPheGlySerSerLeuasnProSerHisaspVallyeHis	20
Db	63512	CGAGCTGATCGCTTATTCTTTTGAACGACCTTGAATCCATCCGATGATGTTAAGCAC	63453
QY	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuYrSer	40
Db	63452	GTAGAGAAGATAAAAAATCGAGTCATGTTCTTGAAGAACTAAGATATTCGGATTGTACTCA	63393
QY	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys	60
Db	63392	GTATGTACAGTTTGAAGAAAAAGCCCTGATTCTGGCTAGGCAAAAATAAGCGCAATAAA	63333
QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
Db	63332	GCTGTAGAGAACTTAACCGCTCTCTCAGCAGCGACTCTCATGTGGGGGGCGTCATAC	63273
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
Db	63272	TTATTGATAAATCGATCATTTTTCACAGCAGTGAATCCAGATTCAGAGTTTTCATTT	63213
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
Db	63212	GAAACAATCTATTATGACGCGCGTGATTTCATGAATTTCTCGTCATATCTTTCCAAAGT	63153
QY	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysIleAlaGlnGlyThr	140
Db	63152	GGAGAAGAAAAATGAAGTCAAGCTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGA	63093
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
Db	63092	TACAGCAGTGATTTCTACTTATTGGTGAAGACCATATTAAGTTTCAGATGAAGAGAGT	63033
QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer	180
Db	63032	CCAAATATATTTTGGTCAAAAGCTTTGGGGGAAAAATCCCTATGTGGAATAACCTTCA	62973
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
Db	62972	GATATCTCCCAAGAAATCGAAAAAGAGTTCAAGTTTATTTGAGGGGTCTGAAGCGAGTCC	62913
QY	201	LysThrGlyAspGlyGlyAsnAlaLysLysLysLysLysLysLysLysLysLysLysLys	220
Db	62912	AAAACTGGCGATGGTGGAAATGCAAGAAGCGAAGAGGCTTCTGATGATGATGATGATG	62853
QY	221	ProArgValThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHis	240
Db	62852	CCCCGGGTCACTGATCCGCCAGTACATGATGATGATGATGATGATGATGATGATGATGAT	62793
QY	241	Met-----	241
Db	62792	ATGGGTAAAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA	62733

QY	242	-----GlyAlaLeuGluSerProlyValIleThrLeuGln	253	Db	61655	CATAATATATGATAGTTGCATACACTAGTGTGTTTAATAATGCATTTTCAGAGTTGG	61596
Db	62732	CTTTGCTGACTATCTTTCTACAGGGGCT- TTGGAGTCCACAAAAGTCAATACACTCCAG	62674	QY	360	l1eAlaLaLeuLeuValIysGlnIleLeuSerHisIysGluSerLeuValArgAlaAen	379
QY	254	SerSerCysLeuSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273	Db	61595	ATTGCAGCCTTGTGGTAAAGCAAAATCTTAGCCACAAAGAATCTCTGGTCGTCGAAAT	61536
Db	62673	TCATCATGTAATCTTTCTCGTACAGATGGTACATNTGGATGGAAATGATGCTTTTGCTTG	62614	QY	380	SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer	399
QY	274	TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293	Db	61535	TCTGAATAGCTTTTCAAAATGCTTAGAGTAGAGTGGATTAATTAATTCGATATTGTCC	61476
Db	62613	TATTTCTATGGCAGCCATATCTCTGGAATCCACAGAGATATGTTAGCTAGTCAAGATTGG	62554	QY	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419
QY	294	GlyLysIleProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetAla	313	Db	61475	TGCTGAAGAGTCTCTTCTGAGCATACACAGGTTTGAGTTGCATGCTTTGGTACT	61416
Db	62553	GGGAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTAAAGCCGAAGTGGCA	62494	QY	420	AnSerLysGlnSerValIleSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr	439
QY	314	LysLeuCysGlnValLeuHisLeuSer-----	322	Db	61415	AATCTAAACAGTCAGTGGTTAGCACAAAACCTAGTAAATGAAAGTCTCTCAGGGGCTACA	61356
Db	62493	AAACTTTGCCAAGTTTTCATCTTTTC- AGTAAGTGGCCTTTTTCACCTCCACAACATTATT	62435	QY	440	ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluCys	459
QY	322	-----	322	Db	61355	GTGCGTCGACAAAAGATTAAATACGAAGTCGATCGGAAATAGCTCAGAGGATGAAGAGTGC	61296
Db	62434	TTAGCCTTGCAATGCTTATATATAGCTGATTGCAACTGTFAGTTGTTACCTGATTTCCCTG	62375	QY	460	MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle	479
QY	322	-----	322	Db	61295	ATGACTGAGAAGAGATGATAGCCATTATAGCACAGCACACAGAGATATCGAAAAGACTATT	61236
Db	62374	TTACAGCCAAATGTGAGAGTTTTATTCTTCACTATATCCATCCGTTTAAGCATATTTTA	62315	QY	480	SerGlyIleLysLysLysGlnValGlnLysLeuValGlnGluHisGluGlu	499
QY	322	-----	322	Db	61235	AGTGGCATAAAAAGAAATACAAAGCAAGTGCAAAAGCTTGTACAAGAGCATGAGGAA	61176
Db	62314	TTTTCTTATCTGGCTTCGTTACCAATGCACCTGTTAAATAGCAACTGCTGCACAAAAC	62255	QY	500	LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys	519
QY	322	-----	322	Db	61175	AAAGAAATGGAGCTGTTAAATATGTATGCAGACAAAGACGAAACTTTGAAACTAGTAAA	61116
Db	62254	AGTAGTAGTATGTGCTCTCATGTCATTCTATTGTTATTGAGCAAGAAATTTCTGTCT	62195	QY	520	SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539
QY	322	-----	322	Db	61115	AGTGGGAAGCAGCAGTAGTAATTCGTATTACCTGTTCCAGGACCAGTACTCAAGTGGTGAT	61056
Db	62194	ACTTTACATGATCCATCTCTGGGAGTATATACTATATATAACCTTAGGCGCTTTGTACCT	62135	QY	540	LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn	559
QY	322	-----	322	Db	61055	CTCAAACTCTGGATCATATATTAATCAAGAAAGTTTGATGAATCAAAAGTCAGAAAAAT	60996
Db	62134	GGCTGATCAAGACATGTCAAAAGTTTATCTGTTGCTGTTGGTATAGAAACTATACAG	62075	QY	560	GluCysLeuLysSerLeuGlnMetHisAspValAlaLysLysLysLeuAlaGluAsp	579
QY	322	-----	322	Db	60995	GAATGCCTCAAAAGTCGAGCAAAATGCACGATGTTGCAAGAAAGAAAGTTGGCTCAGGAT	60936
Db	62074	TGCTGATGCTATTTAAGGTAGTCTTATGCTCTTCACATATTGGCTAATAGATGTTCCG	62015	QY	580	GluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro	599
QY	322	-----	322	Db	60935	GAAGCCTGTTGGATTAAATCGGATAAGAGCTGGCGAGCTAAATTAAGTTTGTGTCTCC	60876
Db	62014	CTGTCGTGCCATATACCTCTGTGATTATCACGGTCTCGTCTATCAAAAATTTGTACTAA	61955	QY	600	IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro	619
QY	323	-----AspAlaCysThr	326	Db	60875	ATTCAAAGTGGCAATAACAAGCATTTTAGTGTTCATCAACATTTCCCAAAATGCTCCT	60816
Db	61954	AAGGTATTTGGCAATGTGATGGTTAAACAGATTATTTGTTTTCAGGATGCTTCACA	61895	QY	620	AspVal 621	
QY	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro	346	Db	60815	GATGTA 60810	
Db	61894	AGCATGGTCGGAATTTTCTCGAATATGTTATTGAAATCACCAGATCTACGAAGGCCA	61835	RESULT 6			
QY	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu-----	357	ATH552638			
Db	61834	GCCACTACTTTTCAGGCATTCAGATAGCCCT-GGTATGACAGCAATTTACTTTGATAAAT	61776	LOCUS			
QY	357	-----	357	DEFINITION			
Db	61775	TATGCATTGTTCCCTTCATCATCTGCGCTTGTGTTTAGAATGCTCAGAGCGAGCACTCC	61716	ACCESSION			
QY	357	-----	357	VERSION			
Db	61715	TTTAGTTTAACTTTCCAAATCATAGGATTCAAATATCCATTAACCTGCGCTTTGATCGCTG	61656	KEYWORDS			
QY	358	-----SerTrp	359	SOURCE			
Db				ORGANISM			
QY				REFERENCE			

ATH552638 695 bp DNA linear PLN 29-MAR-2003
Arabidopsis thaliana T-DNA flanking sequence, left border, clone 340B12.
AJ552638.1 GI:29368785
left border; T-DNA flanking sequence.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

AUTHORS Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M., and Lecharny, A.

TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites

JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)

MEDLINE 22463535

PUBMED 12446565

REFERENCE 2 (bases 1 to 695)

AUTHORS Balzergue, S.

TITLE Direct Submission

JOURNAL Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

FEATURES

source 1..695

LOCUS /organism="Arabidopsis thaliana"

DEFINITION /mol_type="genomic DNA"

ACCESSION /cultivar="Wassilewskija"

VERSION /db_xref="taxon:3702"

KEYWORDS /clone="J40812"

SOURCE /clone_lib="Arabidopsis thaliana T-DNA insertion lines"

ORGANISM 1..695

REFERENCE /note="T-DNA flanking sequence"

misc_feature left border"

ORIGIN

Alignment Scores:

Pred. No.: 1,74e-74 Length: 695

Score: 1033.00 Matches: 203

Percent Similarity: 87.88% Conservative: 0

Best Local Similarity: 87.88% Mismatches: 1

Query Match: 28

DB: 8 Gaps: 1

SEQ3_D1219_COPY_650_1270 (1-621) x ATH552638 (1-695)

QY 104 IleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIleGlyGlyGluGlu 123

DB 2 ATTATGGACGGCGTGAATTCATGAATTCCTGCTCCATCTTTCTTCCAAAGGTGGAAGAA 61

QY 124 AsnGluValIleLeuGlyLeuLeuGluAlaLysHisAlaGlnGlyThrTySerSer 143

DB 62 AATGAAGTCAAGCTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGAACTTACAGCAT 121

QY 144 AspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSerProAnile 163

DB 122 GATTCCTACTTATTTGTTGAGACCATATTAAGTTCTCAGATGAAGAGAGTCCAAATATA 181

QY 164 PheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyPrProSerAspThrPro 183

DB 182 TTTTGTGTCAAAGCTGTGTGGGGGAAATATCTTATGTGGAATATCCCTTCAGATATCTCC 241

QY 184 GlnArgAsnArgLysArgValGlnTyPheGluGlySerGluAlaSerProLysThrGly 203

DB 242 CAAAGGAATCGAAACGAGTTCAGTATTTTGAGGGTTCGAGCGAGTCCCAAAATCTGGC 301

QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223

DB 302 GATGGTGGAAATGCAAGGCGAAGAGAGGCTTCTGTGATGATGTCATCTATCCCCGGTCT 361

QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMet----- 241

Db 362 ACTGATCCGCCAGTAGATGATGATGAAGAAGGGCTCTGGGAAGGATCACATGGGTAA 421

QY 241 ----- 241

Db 422 ATAGTTTAATTTCTGCTCGATACCTCTAGTGTTCATTGATTATGCAACTACTTTCTGTA 481

QY 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256

Db 482 CTATCTTTTCTACAGGGCT-TTGAGTCCACAAAAGTCATAACACTCCAGTCATCATGT 540

QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTySerMet 276

Db 541 AAATCTTCTGGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600

QY 277 GlySerHisIleSerGlyLysProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296

Db 601 GCGACCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 660

QY 297 ProAspGluSerGlnArgLeuHisThrVal 307

Db 661 CCGGATGAATCAGAGGAGGCTCCACACTGTT 693

RESULT 7

AK068790 4027 bp mRNA linear PLN 24-JUL-2003

LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013163J20, full insert sequence.

DEFINITION AK068790

ACCESSION AK068790.1 GI:32978815

VERSION FLI_CDNA; CAP trapper.

KEYWORDS Oryza sativa (japonica cultivar-group)

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1

AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, M., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otsu, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saio, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

TITLE japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 4027)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otsu, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,


```

Db      3995 AAAAAGTTT 4003

RESULT 8
LOCUS   AX063609
DEFINITION Sequence 28 from Patent WO0100801.
ACCESSION AX063609
VERSION   AX063609.1 GI:12541333
KEYWORDS
SOURCE   Brassica oleracea
ORGANISM Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
AUTHORS Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE    Gene involved in epigenetic gene silencing
JOURNAL  Patent: WO 0100801-A 28 04-JAN-2001;
          Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
          m.b.H. (AT)
FEATURES
          source
          1..467
          /organism="Brassica oleracea"
          /mol_type="unassigned DNA"
          /db_xref="taxon:3712"
          /note="seq1-43"
ORIGIN
Alignment Scores:
Pred. No.: 3,82e-31 Length: 467
Score: 494.50 Matches: 97
Percent Similarity: 84.09% Conservative: 14
Best Local Similarity: 73.48% Mismatches: 12
Query Match: 15.37% Indels: 9
DB: 6 Gaps: 2
SEQ3_D1219_COPY_650_1270 (1-621) x AX063609 (1-467)
QY 111 GluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeuCysLeu 130
Db 1 GAATTCCTGTCATATCTTCTTCGATGTTGGAGAGAGAAATGCAAGCAAGCTGTCTTA 60
QY 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTySerSerAspSerThrLeuPheGlyGlu 150
Db 61 CTTTTCGAAGCCCAAGCATGCTCAGGGAAGTTACAGCACTGATCTACTCTATTTCGTGAA 120
QY 151 AspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuGly 170
Db 121 GAACATGTCAAGTTATCAGAT---GAAAGTCCAAATATGTTTGGTCAAGCTGTGAGT 177
QY 171 GlyLysAsnProMetTrpLysTyProSerAspThrProGlnArgAsnArgLysArgVal 190
Db 178 GGAAGAACCCATATGGAATACTGTTGGATCTCTCAAGAGGAGTCGAAAAGAGTA 237
QY 191 GlnTyPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLys 210
Db 238 CGGCATCTTCAGGGCTATGAGGAGACTACCAAGTTGGCAATGGCGAAACTTAAAGAAG 297
QY 211 ArgLysLysAlaSerAspValThrAspProArgValThrAspProValAspAsp 230
Db 298 AAAAAGAAGCTTTCAGATGATGTCACA-----GTAGATAC 333
QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGly 242
Db 334 GCTGAGAGAAAGCCCTCTCGAAGAGATCATCGGT 369
RESULT 9
LOCUS   AX063611
DEFINITION Sequence 30 from Patent WO0100801.
ACCESSION AX063611
VERSION   AX063611.1 GI:12541335
KEYWORDS
SOURCE   Brassica oleracea
ORGANISM Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
AUTHORS Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE    Gene involved in epigenetic gene silencing
JOURNAL  Patent: WO 0100801-A 30 04-JAN-2001;
          Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
          m.b.H. (AT)
FEATURES
          source
          1..515
          /organism="Brassica oleracea"
          /mol_type="unassigned DNA"
          /db_xref="taxon:3712"
          /note="2-33"
ORIGIN
Alignment Scores:
Pred. No.: 5.05e-29 Length: 515
Score: 469.00 Matches: 99
Percent Similarity: 60.87% Conservative: 27
Best Local Similarity: 47.83% Mismatches: 39
Query Match: 14.58% Indels: 42
DB: 6 Gaps: 2
SEQ3_D1219_COPY_650_1270 (1-621) x AX063611 (1-515)
QY 387 SerArgValGluValAspTyIleTySerIleLeuSerCysMetLysSerLeuPheLeu 406
Db 1 TCTAGAGAAAGAGTGGATATATATCTTTCTGCTACTGCAAGAGTCTATTTCGTG 60
QY 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
Db 61 GGGCGCACACAGGTTTCCAA----- 81
QY 427 SerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn 446
Db 81 ----- 81
QY 447 ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSer 466
Db 82 -----GAAAAGGGTGAAGATGTCATGGCTGAGAAAAGAGGTAGC 120
QY 467 HisTySerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysTyTyr 486
Db 121 CATTTAGCTCAGTAACCAAGATGTTGAAAAGACTATTAGGCACATCAAAAAGAATGC 180
QY 487 LysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsn 506
Db 181 AGTAAGAGCCTGCATAAGCTTGTACAAACCCCTCGAGGAGAAAGATGACCTGTGTAAT 240
QY 507 MetTyAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIle 526
Db 241 AGCAATAGCTGTCAAGAGCAGGAACCTTCAGAAATTGTAAAAAGGTGGAAGCATCATTTAT 300
QY 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
Db 301 CGTGTCACCTATTACGTTAATAATCTCAGACCTACATGATGCTCTCCACGGCTGGAA 360
QY 545 HisAsnTyGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSer 564
Db 361 TGTACTTTTGAAGAAGATTTTCATGATCTCAAAAGGAGAGTTGCGATGAATGCCTTGAAGT 420
QY 565 LeuGluGlnMechHisAspValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
Db 421 TTAGAGCAATAAACCAGGCTGGAAGAGAGAGTTGGCTGAAGATGAAGCTGTTGGATT 480
QY 585 AsnArgIleLysSerTrpAla 591
Db 481 AGTCGGATAGAGAAATGGGCA 501

```

```
RESULT 10
AX063607/c
LOCUS      AX063607              519 bp    DNA          linear   PAT 24-JAN-2001
DEFINITION Sequence 26 from Patent WO0100801.
ACCESSION  AX063607
VERSION     AX063607.1  GI:12541331
KEYWORDS   Brassica oleracea
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1
AUTHORS    Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE      Gene involved in epigenetic gene silencing
JOURNAL    Patent: WO 0100801-A 26 04-JAN-2001;
            Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
            m.B.H. (AT)
FEATURES   Location/Qualifiers
            source
            1..519
            /organism="Brassica oleracea"
            /mol_type="unassigned DNA"
            /db_xref="taxon:3712"
            /note="seq1-23"
ORIGIN
Alignment Scores:
Pred. No.:      7,82e-27      Length:      519
Score:           442.00      Matches:     89
Percent Similarity: 84.82%    Conservative: 6
Best Local Similarity: 79.46%  Mismatches:   13
Query Match:     13.74%      Indels:      4
DB:              6           Gaps:        1

SEQ3_D1219_COPY_650_1270 (1-621) x AX063607 (1-519)
QY      1  ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyshHis 20
Db      324 CGCGCTGAGCTCTTTATACCTTTTGGAGCAGCCTGAATCCATCGCATGACGCTTAAGCTC 265
QY      21  ValGluLysIleIysIleGluSerCysSerGluArgThrIlyPheArgIleuTySer 40
Db      264 TTAGAGAAGATAAGGTGAGTGGTGTCTGAAGAACAATGAATTTTCGGTTGTACTCA 205
QY      41  ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
Db      204 GTGTTACAGTTGAGAAAAAACCTGATCTTGCT-----AGGCAAAATAAG 157
QY      61  AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTy 80
Db      156 CCTCTAGAGAACCTAAACCGTCTCTCACACATGCATTGCTCATGTGGGGGGCATCATAT 97
QY      81  LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db      96  TTATTGTATAGCTTGACACACTTCCATGGAGTGGAAACCCAGGTTCCAGAGTTCCATTT 37
QY      101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
Db      36  GAACAGTCTATTAAAGATGCCGTAAANNACGAATTC 1

RESULT 11
AP005621/c
LOCUS      AP005621              184118 bp    DNA          linear   PLN 27-FEB-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
ACCESSION  AP005621
VERSION     AP005621.2  GI:28570002
KEYWORDS   Oryza sativa (japonica cultivar-group)
SOURCE     Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE  1
AUTHORS    Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE      Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
            clone:P0681G01
JOURNAL    Published Only in Database (2002)
REFERENCE  2  (bases 1 to 184118)
AUTHORS    Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE      Direct Submission
JOURNAL    Submitted (01-AUG-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasakienias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT    On Feb 26, 2003 this sequence version replaced gi:22091093.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
FEATURES   Location/Qualifiers
            source
            1..184118
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="genomic DNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /chromosome="6"
            /clone="P0681G01"
ORIGIN
Alignment Scores:
Pred. No.:      1.02e-19      Length:      184118
Score:           397.00      Matches:     178
Percent Similarity: 32.47%    Conservative: 134
Best Local Similarity: 18.52%  Mismatches:   255
Query Match:     12.34%      Indels:     395
DB:              8           Gaps:        22

SEQ3_D1219_COPY_650_1270 (1-621) x AP005621 (1-184118)
QY      3  AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyshHisValGlu 22
Db      19304 GATGCATCATCATATATTGTAGTGATGGAAATCCAACCAATGATCGAGAGTTCTCCAG 19245
QY      23  LysIleLysIleGluSerCysSerGluArgThrIlyPheArgIleuTySerValCys 42
Db      19244 AGGATCAGTATAGAGTCTCAATCTGAATGTGTGCTATTTTCGTTGTATTCGTCGTGT 19185
QY      43  ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db      19184 ACGGTGGAGGAAAAGACTCTTATATAGCAAGCATGATCATATTCTTCACAGCAACGTC 19125
QY      63  GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyIleuPhe 82
Db      19124 CAGATGTAATGCCCTATCGTAGCCATTCCTGTAGTTGGGTGGCATCGTTCTCTTC 19065
QY      83  AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
Db      19064 AATAGACTTGAGGAGTTCAAAAGCATGACTACTCCAGTAAAGATCTCGAGATGATGTT 19005
QY      98  ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
Db      19004 CTC-----TTCATGAATAATGTATTTTGGAGTTTGCAGCGCAACATATCC 18960
QY      118 SerLysGlyGlyGluGluAsnGluVallyshLysLeuLeuLeuAlaLysHisAla 137
Db      18959 ACCAACGTTGAAGCCAGCACTAAATGAAATGCAATATCTCGAGCTCAGCAGAGT 18900
QY      138 GlnGlyThrTySerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSer 156
Db      18899 GGATCGTTTTATTCTAGAGATATTGCTGTATTAACGAGAGGGAGGAATATCTCAGTT 18840
QY      157 AspGluGluSerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
Db      18839 GATGCGTATTTGCCAAAGTTCTGGACCTTCTGGTCAAAATTTTACTAGTGGGAGGCTCTCT 18780
QY      175 MetTrpLysTySerProSerAspThrProGlnArgAsnArgLysArgValGlnTyPheGlu 194
```

```
Db 18779 CATTGGCAATATATCTGAGCCAGTGCAGAACCGTAGGAAGATACAAATATGGAA 18720
Qy 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAla 214
Db 18719 GACCAATGAGATACCTGTCAGAGAACTGTAGTGAAGCAATATGAACGTAGAAAATC 18660
Qy 215 SerAspAspVal---ThrAspProArgValThrAspProValAspAspGluArg 233
Db 18659 GGAGAAATCATGATTCATCTCCGAAGATCCCTA-----CCTGTCAAAGACATGAT--- 18609
Qy 234 LysAlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 18608 -----CCTGTGTCGCTCGAA 18594
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGly----- 267
Db 18593 AATAGTACAGCATCCAGTCTCATGAACACATCAGTTGTAGTGTGACACTTGGCAAGAGCTAGT 18534
Qy 268 -----AsnAspAlaPheGlyLeuTyzSerMetGlySerHisIle 280
Db 18533 ATCCAACTGTAAACAGTAATTAACAGCTCTTATGCTTTATGTGCATATGCTACATTTT 18474
Qy 281 SerGlyIleProGluAspMetLeuAlaSerGlnAspTyrGlyLysIleProAspGluSer 300
Db 18473 TATAATACACTCCATTTGTAGGGCGAGAAAGCCTTCAGGCG-----ACA 18429
Qy 301 GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis 320
Db 18428 CAGAAAGGCTTTCACACACCACTCAAGCCAGAGCTGTCAAAGTTGTATGAATTTGTGGAA 18369
Qy 321 LeuSer----- 322
Db 18368 TTGCGGCTACTGCCAAGTATATTAAATAAATGCTTGTCTTTATGTTATTAAATTTGAGAAA 18309
Qy 322 ----- 322
Db 18308 AATATTTTAGTTCACGCAAGTACAGGACATAATCTGCTCTTTTAGCATTTCAATTTAATA 18249
Qy 322 ----- 322
Db 18248 GCTAATATTTTGTAGTTTTCACAGCAGAGATCCACACCTGTGAGAGATTATTATCAA 18189
Qy 322 ----- 322
Db 18188 GTTGGATCTTGATACTTTTGTCTTTCATAACGGAACATTTTCTTTGATCTCTAGTGTG 18129
Qy 322 ----- 322
Db 18128 TCATTTTGATCCATATATGTTACTTCTCCGTTTCAAGTAAAGACTTTCTAGCATTCG 18069
Qy 322 ----- 322
Db 18068 CCACATTCATTAGATTGTAATGAATCCAGACATATGTATGTGTTTAGATTCAATTAACAT 18009
Qy 322 ----- 322
Db 18008 ATATATGTATGTGGCAATACTAGAAAGTCTTACATTTGTGAATGAGGAGATTAGTT 17949
Qy 322 ----- 322
Db 17948 GCTACTATATCCGCTTTTAGCTCAGGTTATAGCTTTTGATATCACCTACTCCCTCCGCC 17889
Qy 323 -----AspAlaCysThrSer----- 327
Db 17888 CCATAGAAAACCACTAGTACTGGATGCTGTCTAGATTCTGTAGTACTAGGATGTGTCA 17829
Qy 327 ----- 327
Db 17828 CATCCGCTACTAGTTGGTTTCTATGGGACGGAGGTGCATGTTTCTCATGCTCTCC 17769
Qy 327 ----- 327
Db 17768 AATATGAATTTATGCAGGACAGCTGTGTCCTGCTGCTGCTGTTTAAAGTTGCTCCTCT 17709
```

```
Qy 327 ----- 327
Db 17708 CTTCTGTAGAAATTTATGCAGCATGCTGTTTACCCTAAGTTTTCAGTTTAAAGCAACTTTTCC 17649
Qy 327 ----- 327
Db 17648 TGAAGTCACTGATTTCTTAGTTTAAATATAATATAAGCTGATGATGCTGCTGGATGCTTGTAT 17589
Qy 328 -----MetValGlyAsnPheLeuGluTyrValIle 337
Db 17588 AAAACCTTTTGCAGGAGACTGTCAAATGTTTATGTGAGGAATTACTTGAT-TACATTGTTG 17530
Qy 338 GluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeu 357
Db 17529 AAGAATCATCAAGTCAAGTCAAGGCAAGGCGCATATTTCATGTCATTCACATTCGCCCTG 17470
Qy 357 ----- 357
Db 17469 GTTAGATTATAAAATATTTTATAGTTTTTTAGCTTAAATTTGTCATGGCTTAAATGCTATT 17410
Qy 358 -----SerTyrIleAlaAlaLeuValLysGlnIleLeuSe 370
Db 17409 ATTTTTCATGCTGCTTTTTCAGTGTGGGTGCTGCTCTCTCTTAAAGCATAAAGATCAA 17350
Qy 370 rHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGI 390
Db 17349 TCGAAGAGAGTCACTTGCCTTTCTGTGAGAAACTTGAACACTATGAGTGGCATGAAGTCT 17290
Qy 390 uValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThr-- 409
Db 17289 CGCTGAATATGTTTATGAGAAGCTAAGGATCCTCAAGAAAAAGTTTTCGCGTAGACAAG 17230
Qy 410 -----GlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerVa 425
Db 17229 TGAACAACGACAGAGCAATCTACTCCAGTAATAACATCATCTTACAAACAACA 17170
Qy 425 lValSerThrLysLeuValAsnGluSer-----LeuSerGlyAlaThrValArgAspGI 443
Db 17169 GACCTCAGCAATTTGAGAAGTGAAGGATCAATTTGTCTATCAGGTAACGACATTCATGG 17110
Qy 443 uLysIleAsnThrLysSerMetArgAsnSerSerGluAsp-----GluGluCysMe 460
Db 17109 TCATTGGAAAAATGTTTTCATGAGAGGCTCCACATGATATCTTTCGACCGAGAGATGAT 17050
Qy 460 tThrGluLysArgCysSerHisTyz-SerThrAlaThrArgAspIleGluLysThrIleSe 480
Db 17049 ATTAGAACAGAG-----GNAATTGATATC 17026
Qy 480 rGlyIleLysLysLysTyrLysLysGlnVal-----GlnLysLeuValGlnGluHisGI 498
Db 17025 TGTCTTAGAACTCATAGAGAAACATGTTTAAAGGGATGAACCTTCTTGAAGAATCAC 16966
Qy 498 uGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLys----- 514
Db 16965 AGAGAAAAGAATTAATTTAAATGATGTTTCTCTCTTAAAGAGAAAGAAATATTTCAGA 16906
Qy 515 -----LeuGluThrSerLysSerValGluAlaAlaValI1 526
Db 16905 TAAACAGGAACAGACACACTGTTAGCATGTCACAGCAAGAAAGAGTGCAGAGCT 16846
Qy 526 eArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAs 546
Db 16845 GCAAGAAACATCAAT-----TTGGTCTGTGACATCT 16813
Qy 546 nTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGI 566
Db 16812 TCGTAAAGGTCACTTGTAT-----TCAGAGACAGGGATGCCCGGTGAAGCTAATAAT 16759
Qy 566 uGlnMetHisAspValAlaLysLysLeuAlaGluAspGluAlaCysTyrIleAsnAr 586
Db 16758 TGAATGTTTACTTCTGCTGATGCTATTTTGAATTCATGAGATGCAGCATTAACA 16699
```

```

QY 586 gilelys-----SerTrpAlaLalysLeuLysValCys---ValPro1 600
   ::::|
   ::::|
Db 16698 GTTGAAGATGACGAATCACTTCATGGAACAAGGAGTTACAATTGAAGGAAATTTTCCT 16639
   ::::|
QY 600 eGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsn1LeSerGlnAsnAlaProAs 620
   ::::|
Db 16638 TCAGCAAGCAAAATCTGGCCACTTAGATCGTAGCTTTGTAGTCAACAAATTCCTTTACCAGA 16579
   ::::|
QY 620 p 620
   ::::|
Db 16578 T 16578

RESULT 12
LOCUS AP005005
DEFINITION Orzya sativa (japonica cultivar-group) chromosome 2 clone P0513D05,
*** SEQUENCING IN PROGRESS ***.
ACCESSION AP005005
VERSION AP005005.1 GI:19773545
KEYWORDS HTG; HTGS PHASE2.
SOURCE Orzya sativa (japonica cultivar-group)
ORGANISM Orzya sativa (japonica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Orzya sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
       clone:P0513D05
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 150597)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Takuji Sasaki, National Institute of
       Agrobiological Sciences, Rice Genome Research Program; Kannondai
       2-1-2, Tsukuba, Ibaraki 305-8602, Japan
       (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
       Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
       are represented as runs of N. The order of the pieces is believed
       to be correct as given, however the sizes of the gaps between them
       are based on estimates that have provided by the submitter. This
       sequence will be replaced by the finished sequence as soon as it is
       available and the accession number will be preserved.
       * NOTE: This is a 'working draft' sequence.
       * This sequence will be replaced
       * by the finished sequence as soon as it is available and
       * the accession number will be preserved.
FEATURES
       Location/Qualifiers
       1..150597
       /organism="Orzya sativa (japonica cultivar-group)"
       /mol_type="genomic DNA"
       /cultivar="Nipponbare"
       /db_xref="taxon:39947"
       /chromosomes="2"
       /clone="P0513D05"
ORIGIN
Alignment Scores:
Pred. No.: 1,29e-14 Length: 150597
Score: 332.50 Matches: 199
Percent Similarity: 30.64% Conservative: 108
Best Local Similarity: 19.86% Mismatches: 269
Query Match: 10.34% Indels: 428
DB: 2 Gaps: 28

SEQ3_D1219_COPY_650_1270 (1-621) x AP005005 (1-150597)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
   ::::|
   ::::|
Db 137548 GATGCCATCATATATGTTCTTGACAAACACCACCTAAACGACTTGAAGGCTCTTCAA 137607
   ::::|
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
   ::::|
   ::::|

```

```
Db 138685 TTATCCTATTACATGCTGTGAATGACTATATCAATATATTTTGAAGAAATACATACAGTTT 138744
Qy 277 ---GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp----- 293
Db 138745 AATGGAGCGAAATAGCAGGA-----CAAGTTCTCATGGA-AGTAGTGAGTGATGTTG 138797
Qy 293 ----- 293
Db 138798 TTGTTTCTACTAACTATCATTTTCCCAAGGTGGGAAATAAATAGTGTGGAATCT 138857
Qy 293 ----- 293
Db 138858 TTCTTTGATCTTTACTAAATAATTGGAAGTTTCATGATGGTGACGGTACACTCATCCAGCT 138917
Qy 293 ----- 293
Db 138918 TRAAGGTAGCATCAATCATATAAAGAGTAGCGCTATCAACCTAGATTAAAGCAGCGATT 138977
Qy 293 ----- 293
Db 138978 CTCACAACCTAAATCCTTAAGATATTCAACCATGAAAAATTTGTAGAGGTGCAGTTATTTC 139037
Qy 294 -----GlyIleProAspGluSerGlnArgArg 303
Db 139038 ATGGACTTAGCTCATTATGCTATAAGTGACAAAAGCCCTCAAACTCAGTCAAGAGA 139097
Qy 304 Leu----- 304
Db 139098 TTACGGAGCGAAAGTCCACTGTAAAGGAGCTCTAACTATTAAACCTTGAATCATGAGA 139157
Qy 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis----- 320
Db 139158 CATTTAGGGGAAAAAACCCTGAAGTAGAAAAGCTGGCTGT-GTCTGACACTCCAGCTGTCCG 139216
Qy 320 ----- 320
Db 139217 CAGTTCAGTCACACTTGTCTTAAATGGGAAAAAATAATCTTGTGTAAGTCTTTTTTTGG 139276
Qy 320 ----- 320
Db 139277 TTTTATCATATAACTCTTACCTTTGTTAAAGCTGATCTCATATTATTTGTAATTTTCTGTT 139336
Qy 321 -----LeuSerAspAlaCys 325
Db 139337 GCTTTGGCTTGTATCACCACCTAAATGCTTTTCGCAATTAATTAATTTGCAGGACAATGTT 139396
Qy 326 ThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu 345
Db 139397 ACCTTGGTGGTGAGCAGTTCTTTGAATATCTTTTGAATAATCATGTAGTTGTTTCAGGAG 139456
Qy 346 ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla----- 361
Db 139457 CCAAAGTACATATTTTCATGGTTGAACATAGCTTTTGGTACGATAACCCCAATATC 139516
Qy 362 -----Ala-LeuLeuValLysGlnI 368
Db 139517 GTCGTTCTTTTAAATTTTCAGATTTCATATTGACCAAAATGCTAATGCTGTAATACTTT 139576
Qy 368 eLeu-----SerHisLysGluSerLe 375
Db 139577 CCTTCAGTGTGCGAGTTGCCCTCTATTATCAACTTTTAAGGTGGATCAATAAGAGTCACCT 139636
Qy 375 uValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTy 395
Db 139637 TGCCCTTGTGAAAAAGGTTAAATACGATGTCATGAGGAGCTGCAGAGTTGTTGCTTA 139696
Qy 395 rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs 415
Db 139697 TGACAGCCTA-----AAGAGAAAAATTTCCAAAGAAAGCAGGTGCAACTGGCAGGCAA 139747
Qy 415 pCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLe 435
Db 139748 TTGTCAATCAACA-----TCAGTAGAAAAAGACCAAAACCTTTCAGCAGGAGGAC 139795
```

```
Qy 435 uSerGlyAlaThrValArgAspGluLysIleAsnThrLys----- 448
Db 139796 CTCTAAT--ATCTTGAGAAATGACCACATTTTCCCTAAGCAGAGAATGGATCTTCATGA 139852
Qy 449 -----SerMetArgAsnSerSerG 455
Db 139853 TAATTTTCATGAATGGTGCACTTCAAGAAGCGCTCATTTGTTGTCGCAAAATGGTTTCAGA 139912
Qy 455 uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyrSerThrAl 471
Db 139913 AGAGCAGGAAGTATGATGATGTTTCCAGGAACCTCATGGAATGC---CATTTTCAACTGA 139969
Qy 471 aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysLysTyrLysG 489
Db 139970 TGAACCTTCTCGCATTTGTGGAGAAAGAAATCAACTTAATTTGACAATGTTTCTCTGTAAG 140029
Qy 489 nValGlnLysLeuValGlnGluHisGluLysLysMetGluLeuLeuAsnMetTyrAl 509
Db 140030 AGAATACCGTATCTTTGTGATAAACACAGCTCTCAGATTTCAGAGTTAGAGAAATACACACA 140089
Qy 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleTh 529
Db 140090 AAATAAAACTGCCAGACTGAAAACA-----GT 140116
Qy 529 rCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluAr 549
Db 140117 GTGCAAT-----TTAGTTCTGGAACATATTTGTAGAAG 140149
Qy 549 glySerPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHi 569
Db 140150 TCATGCTGAT-----GTAGACACCGNAATGACACANATAAAACAAACTGTTCAAGTGGTT 140203
Qy 569 sAspValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSe 589
Db 140204 CACTATGCTGATGATGCTATTCCTGAGGCACATGAGACTCCAGCAGCAAACTTGAGTC 140263
Qy 589 r-----TrpAla-----AlaLysLeuLysValCysValProI 600
Db 140264 ACTGCAATCTAATACATGCGCTGAGAGGAGCGACAGCTGAGGAAAAAACTTGTGTTGAAGC 140323
Qy 600 eGlnSerGly----- 603
Db 140324 AAAATCAGGCCAATTAGATCATCACCTTTGATCAGCAGATTGCTCTACCAAGATTCAAACTT 140383
Qy 604 -----AsnAsnLysHisPheSerGlySe 611
Db 140384 TGTATGCAAGAATTTCATCCATTTGAGGACAAAGTAGACNATCCCATGCTCAGTGGAG 140443
Qy 611 rSer 612
Db 140444 CGCT 140447
RESULT 13
AP004752 156603 bp DNA linear PLN 21-JAN-2004
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
DEFINITION PAC clone: P0501G04.
ACCESSION AP004752
VERSION AP004752.3 GI:41052917
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
clone: P0501G04
JOURNAL Published Only in Database (2002)
REFERENCE
2 (bases 1 to 156603)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
```

TITLE Direct Submission
JOURNAL Submitted (13-FEB-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7411, Fax: 81-298-38-7468)
COMMENT On Jan 21, 2004 this sequence version replaced gi:34740248.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://opal.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM
(http://www.tigr.org/cdb/glimmer/glmr_form.html), RiceHMM
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.lastate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/docs/sim4.html), gap2
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI NonRedundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DDBJ. Protein homologies of the coding
regions were searched against NCBI NonRedundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
This sequence of P0501G04 clone has an overlap with P0017C12 (DDBJ:
AP004769) clone at 5' end and with OJ1399.H05 (DDBJ: AP004090)
clone at 3' end. Detailed information on overlap and assembly
quality together with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/Genomeseq.html.
FEATURES
Location/Qualifiers
1..156603
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="P0501G04"
complement(join(2875..3454,3805..3815))
/gene="P0501G04.1"
complement(join(2875..3454,3805..3815))
/gene="P0501G04.1"
/note="gag-pol-like protein"
complement(4646..7455)
/gene="P0501G04.2"
complement(join(<4646..4662,5391..5429,6120..6360,
7423..7455))
/gene="P0501G04.2"
/note="start and end point are not identified"
complement(join(4646..4662,5391..5429,6120..6360,
7423..7455))
/gene="P0501G04.2"
/note="predicted by RiceHMM etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07829.1"
/db_xref="GI:41052918"
/translation="MKRLYSHKSKIGSDVDKENVVFKPIERSFLDMLDSSVSGADLTQ
DKNKINELLIIHQFGDNCIDERAKNLLVDVFLVLSNSKPIVPENTTVVDGIMKKLS
KPDYVSA"
10034..10234
/gene="P0501G04.3"

misc_feature

10034..10234
/gene="P0501G04.3"
/note="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
complement(113016..16261)
/gene="P0501G04.4"
complement(join(13016..13562,14371..15279,15556..16261))
/gene="P0501G04.4"
complement(join(13328..13562,14371..15279,15556..16211))
/note="supported by full-length cDNA (s): AK069014"
/gene="P0501G04.4"
complement(join(13328..13562,14371..15279,15556..16211))
/note="contains EST(s): AU031908(R2510),D28321(R2510)
contains full-length cDNA(s): AK069014"
/codon_start=1
/product="kinesin light chain-like protein"
/protein_id="BAD07830.1"
/db_xref="GI:41052919"
/translation="MPGLAADNSPPVAAPPRLRLSSPLPRSRAPSPSPSTSSRAKP
RKVAAPETDAALNDPGLPFLIKQADWVGEGGAARALEFAERAAERARERGE
GAIEELAMSLHVAADIHCLGRHADAIVLERAVAVVTPPPPPPESSAEQFPE
DQKGEHWALAAFQSMQMGTHAMLGMDSEIACYKGLSEIQMAALGERDRAVETC
RYLAEAHVQALQFDEAEKLCRKALIEHREHSPASLEASDRRLMALILDAKGDVGA
LEHLVLSMTWANGRDIEVATIDVAICNTYLAARDEAVFSYKALTVLKASRGDD
HPSVASVFRILADLYHRTGRLESSEYCNALRYVAKPAPGAPEVAGGLMEIAIY
EALGDLBALKQLALKLLSDSGQWSTVAGIEAQMGVLYTMVGRYADGRNSFNAY
AKLRASGRKSAFFGVLLNQMLACVQLFKIDEAAQLFEARAVLEQCGASHPTLG
VYSNLAAIYDAMGRVEDAIEILHVLKVEEKLGTANPDVEDEKLRLAELLKEAGRSR
NRKOKSLBNLFTVNSQVRVKDAGRWSNFGFRS"
18235..22976
/gene="P0501G04.5"
join(<18235..18636,19319..19528,19610..20071,20165..20275,
20362..20541,21136..21420,22634..22754,22855..>22976)
/gene="P0501G04.5"
/note="start and end point are not identified"
join(18235..18636,19319..19528,19610..20071,20165..20275,
20362..20541,21136..21420,22634..22754,22855..22976)
/gene="P0501G04.5"
/note="similar to Arabidopsis thaliana chromosome 1,
Fik23.23"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07831.1"
/db_xref="GI:41052920"
/translation="MPSPLLHHHHLSPRFPSPSPRTLSLHPAPRLRLRL
RASARGGGSTRDHDGARSIPARCYEAALRLDSGTARREQVAJAAAAADGGAA
EAHLAAVADAMVEAFPGPDGARSAMSTRILIQANVTEKAKLKDGLTFFESSE
EPDAESVLAMAFQVVMQDLNFRLEVPSEI DLOLSKPKPKVPADPFSRSDEKL
LAALAEAFPCVIEDARNYLGCTGLFHKMKNSCLSDSSVICDIIISEVSVNSARR
LDSFOLVSSHVAGKAKGWPAPASERLVKIGDGFMLMASEFYTYKLQIDAKAPK
NTKLGHHVLANNKGEVLLSHAQMVLELANILDMYFDQTLPGKTFYGNWNSPKIK
KNNGYINNLFALLAGSSIVFLVGAIAQLCWPSLKDRLATVSSCVSSQSYCSDIHS
LDSSELOAYCVSVVEKIKDSFGCTGLDMVANDIGAWGELPFFKGINCDHDDVDI
QSTGAI SQGEQPSLVSPDKMSHLEQNDTQETLQNTASFOVNSRGKVGVQFTN
RLAVNHWATNPLTKLLYSGRKISPAFLERLRISRPKAVVPVELLSVNPSEFFALVR
PVQDPC"
complement(23753..26108)
/gene="P0501G04.6"
complement(join(23753..24276,24367..24619,24688..24918,
25218..25348,25506..26108))
/gene="P0501G04.6"
/note="supported by full-length cDNA (s): AK106946"
complement(join(24086..24276,24367..24619,24688..24918,
25218..25348,25506..26061))
/gene="P0501G04.6"
/note="contains full-length cDNA (s): AK106946"
/codon_start=1
/product="putative GDSL-motif lipase/hydrolase protein"
/protein_id="BAD07832.1"
/db_xref="GI:41052921"
/translation="NRRLLAVAAALLLVAMHFLSLITMAAQLSPFPQDDQYDDPPMP
GLPVSPSPGYSDSPEPLPDSPPEPAPPTPQQQQQWSPQPPRPPFAP
PRTVVPQPGWSSVPPPPARVINVTGCTTMTLVFGDSTVDPGNNRLQTAKANFL

PYGADFLGGRPTGRFNGRLITDILAEKLGIAIRSIQFRDPRRLRSQQLRRGVSPASAG
 SGYDEATRSNALSFPNQILSRVANYTQVMAALGRRFVFGVPPICGLPFIARTLL
 FYLAAGQAAEAGPGYQNLISRVANYTQVMAALGRRFVFGVPPICGLPFIARTLL
 GTTTRCHENMLLATSFNFRILVEVRLKLNQINRATFVDVITTTGMATISPNNYGL
 TETSRGCCGTGVIEVGTQCRGRACHTPSKYIYWDAAHHTERMQNIITEVIMNSIGE
 IVA"
 complement (27678..29619)
 /gene="P0501G04.7"
 complement (join<27678..27819,28014..28543,28645..28787,
 29070..>29619)
 /gene="P0501G04.7"
 /notes="start and end point are not identified"
 complement (join(27678..27819,28014..28543,28645..28787,
 29070..29619))
 /gene="P0501G04.7"
 /codon_start=1
 /product="putative zinc-finger protein"
 /protein_id="BAD07833.1"
 /db_xref="GI:41052922"
 /translation="MRVLITTYVIVVVVTHVMDAGDGGGGVAVVPSCDSRGG
 ERAVHCAQHGGRCLVCDLAHAAARARHRAPLCDACHAAPAVARSQDHAEAF
 ASCARDASRCHRPAAAYTGPGEAMARIFSIDVAPPPPPAAATDEPWTIDLPF
 VPDNSMSTNYCYEPEVLKEANDANPDISKEMATIGREDLLVDNANQDYFOAWTNS
 FDSVALMEPCALQERSYLDLPSYFDLSYLPDHOQMASSSCSDIALLSDFLOPL
 NMSNAYVQLPMMDANINNEIGATSSSLAQILPOSSDHSLLQPLINDETAIDQPL
 VIDTNSNNNTGSEFFCVNFSQNTSGLLGSSNMFQDQQTSHIVLPEKCPDEK
 RQRAVQRYEKKSNRRFVKQIMYASRKATADTERRVRGRFVKASLEQGTSSNDNKQPK
 HEGN"
 complement (join(31481..31619,31721..31782))
 /gene="P0501G04.8"
 complement (join(31481..31619,31721..31782))
 /gene="P0501G04.8"
 /notes="hypothetical ORF
 predicted by GENSCAN
 this category is not included in IRGSP standard"
 32632..34373

Alignment Scores:
 Pred. No.: 1.36e-14 Length: 156603
 Score: 332.50 Matches: 199
 Percent Similarity: 30.64% Conservative: 108
 Best Local Similarity: 19.86% Mismatches: 269
 Query Match: 10.34% Indels: 428
 DB: 8 Gaps: 28

SEQ3_D1219_COPY_650_1270 (1-621) x AP004752 (1-156603)

QY	3	AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValIysHisValGlu 22	
DB	59263	GATGCCATCATCATATATGTTCTGACAAACCAACCACTAAACGACTTGAAGGCTCTTCAA 59204	
QY	23	LysIleLeuValGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42	
DB	59203	AGATCAAGATAGTACACATTTGAGCGGTGAGCATTTTTCGTTGATACATCTCTTTT 59144	
QY	43	ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62	
DB	59143	ACAGTTGAGGAGAAGAGTCTGTCTTCGAGGCGAGGCATTTGTTTATGATAACAACATA 59084	
QY	63	GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82	
DB	59083	CAAGATCTAAGGACTAGTTTAAACATTCCTTACTCAGATGGGGTGTGTCATCTCTCTTC 59024	
QY	83	AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102	
DB	59023	AGTAGACTTGATGAGTCCAGCAAGATGACCATGCAAGTAAAGTTCTGCAATGGGAGAGA 58964	
QY	103	SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlyGlu 122	
DB	58963	CACTTTATTGATCAAGTATCGTTGAGTCTTAAACAAAGTTATCCACACTGTTGAGAC 58904	
QY	123	GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142	

DB	58903	AGCACTGAAGTACACAGGAAGTCCATAGCAAAAGCTAAATATGAGTGGGNACTATATTCA 58844
QY	143	SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
DB	58843	AGAAATATTACTCTGATGGTGAAGAGGAGGAATCTCTGTGCTTGAGGATAATCCAGCT 58784
QY	163	IlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAspThr 182
DB	58783	GAGTCTCTGTTAAATTTATTGGATGGAAGATCCCTCATGTGTGAGCTGTATATCTGAGCCA 58724
QY	183	ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202
DB	58723	CTGCAATCAAGATTACAAAGTCCAGACCATGATGAGTCAATGCTCGGCTGAAGAA 58664
QY	203	GlyAspGlyGlyAsnAlaLysLysArgLys- 212
DB	58663	ATCAATGAAGCCAGAAAGAGCATAGAAAGTTGGTGAGATTATGGTTTCATCTTCAAAA 58604
QY	213	-----LysAlaSerAspValThrAspProArgValThrAsp- 225
DB	58603	GTTGTATCTGCACAAAAGTAAACGATGATGATTCGCTCGATATTTGTACTACATCTGCGCCT 58544
QY	226	-----ProProValAspAspGluArgLysAlaSerGly- 237
DB	58543	GCTCTTCAACCACTTGTGTCACGACGACAAA-TCAGGTATCACTATTTAATAATAA 58487
QY	237	----- 237
DB	58486	TAATAATATTTTCATGTGCTGTGAATATGCTATAGCATTTTTTGTCTCTTTTCTTTG 58427
QY	238	LysAspHisMetGly- 249
DB	58426	CGGCTGCACATTCGCACTTTTCTAACACAGTTCATCTGAAGGTCGGAAGCCTGATGAG 58367
QY	250	IleThrLeuGlnSerSerCys- 256
DB	58366	TACACCCCAAGATCTTCATGCCCAATGAAGCAGAGTGTGCGAAATTAATTAAGGTGT 58307
QY	257	LysSerSerGlyThrAspGlyThrLeuAspGlyAsn- 268
DB	58306	ACAACCTACCGGTACGGATATCCTAAGCTCTGTCAATTTTTTTCCTTATGATACACTTTA 58247
QY	268	----- 268
DB	58246	GCATCATGCGCTGAAGTCATATCTATCTCTATCATATCTTTTGAAGCATGATGTACTA 58187
QY	269	-----AspAlaPheGly- 272
DB	58186	GTTAGTGCTTATGCAATTCCTATGCGCTACGGGTTTCAATAAACAAGAAATAACATACA 58127
QY	273	-----LeuTyrSerMet- 276
DB	58126	TTATCCTATTACATGGTGTAAATGACTATATCAATATATTTGAAGATAACAATACAGTTT 58067
QY	277	---GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp- 293
DB	58066	AATGGAGCGAAAATAGCAGGA-----CAAGTCTCATGGA-AGTAGTGAGTGGATGTTG 58014
QY	293	----- 293
DB	58013	TTGTTTCTTACTAAACTATCATTTTCCCAAGGTGGGAAATAATAGTGTGGAACTCT 57954
QY	293	----- 293
DB	57953	TTCTTTGATCTTTACTAAANATTTGAAGTTCATGATGGTGACGGTACACTCATCCAGCT 57894
QY	293	----- 293
DB	57893	TAAGATGATCATCAATCATATAAAGAGTAGCGCTATCAACCTAGATATTAAAGACGCGATT 57834
QY	293	----- 293
DB	57833	CTCCACACTAAATCCTTTAAGATATTTCAACCATGAAAAATTTGTAGAGGTGCGAGTTATTC 57774

QY	294	-----GlyLysIleProAspGluSerGlnArgArg	303	QY	509	aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleTh	529
Db	57773	ATGGACTTAGCTCCTTATGCTATAAGTGAAACAAAGAGCCCTCAAAACTCAGTCAAGAGA	57714	Db	56721	AAATAAAACTGCACAGACTGAAACA	56695
QY	304	Leu-----	304	QY	529	rCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluAr	549
Db	57713	TTACGGAGGCAAAAGTCCACTGTAAAGGAGCTCTAACTATTAAACACCTTGAATCATGAGA	57654	Db	56694	GTGCAAT-----TTAGTCTCGGAACATATTTTGTAGAG	56662
QY	305	HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis-----	320	QY	549	glpysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMethi	569
Db	57653	CATTAGGGGAAAAAACCCTGAGTAGAAGAGCTGGCTGT--GTCTGACACTCCAGCTGTCGG	57595	Db	56661	TCATCTCTGAT-----GTAGACCCGGAATGCACACATATAACAACTGTTCACTGGTT	56608
QY	320	-----	320	QY	569	aAspValalalysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSe	589
Db	57594	CAGTTTCAGTCACACTTGTCTCTTAATGGGAAAAATAATCTTGTGTAAGTCTTTTTTTGG	57535	Db	56607	CACATATGCTGATGTATGATTCCTGAGCAGCATGAGACTCCAGCAGCAGCAAACTTGAGTC	56548
QY	320	-----	320	QY	589	r-----TrpAla-----AlaLysLeuLysValCysValProI	600
Db	57534	TTTTATCATATAACTTTTACCTTTTGTATAAGCTGATCTCATTTATTGTATAATTTTCTGTT	57475	Db	56547	ACTGCAATCTATATACATGGCTGGAAGCGGACAGCTGGAAGGAAAACCTTTGTTGGAAGC	56488
QY	321	-----LeuSerAspAlaCys	325	QY	600	eGlnSerGly-----	603
Db	57474	GCTTTGGCTTGTATCATCACACCTAATGCTTTTCGATTAATTAATTAATGAGCAATGTT	57415	Db	56487	AAATCAGGCCCAATTAGATCATCACCTTTGATCAGCAGATGCTCTACAGATTTCAAATTT	56428
QY	326	ThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu	345	QY	604	-----AsnAsnLysHisPheSerGlySe	611
Db	57414	ACGTTGCTGTTGAGCAGTCTTTTGAATATCTTTTGAATAATCATGCTAGTTGTTCAAGGAG	57355	Db	56427	TGTTATGCAAGAATTCATCCATTTTGAAGGAAACAAAGTAGCAATTCCTCATGTGAGTGAAG	56368
QY	346	ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla-----	361	QY	611	rSer	612
Db	57354	CCAAAGTACATATTTTCATGCTGTGAACATAGCTTTTGTGTAGCAATAACCAACCCCAATATC	57295	Db	56367	CGCT	56364
QY	362	-----Ala-LeuLeuLysGlnI	368	RESULT 14			
Db	57294	GTCGTTCTTTTAAATTTTCAGATTCTATATTGACCAAAATGCTAATGCTCGTAAATACTTT	57235	LOCUS	BC046638		
QY	368	eLeu-----SerHisLysGluSerLe	375	DEFINITION	BC046638.1	GI:28436850	
Db	57234	CCTTCAGTGTGGCGAGTGGCTCTATTATCAACTTTTAAGGTGGATCATATAAGAGTCACT	57175	ACCESSION	BC046638		
QY	375	uValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTy	395	VERSION	BC046638.1		
Db	57174	TGCCCTTGTGAAAAAGGTTTAAATACAGTGCATGAGGAGCTGGCAAGGTGCTCTA	57115	KEYWORDS			
QY	395	rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs	415	SOURCE			
Db	57114	TGACAGCCTA-----AAGAGAAAATTTCAAGAAAGAGAGCTGGCACTGGCGAGCA	57064	ORGANISM			
QY	415	pCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLe	435	REFERENCE			
Db	57063	TGTCAATCAACA-----TCAGTAGAAAAGACCAAACTTCACAGCAGGAGAC	57016	AUTHORS			
QY	435	uSerGlyAlaThrValArgAspGluLysIleAsnThrLys-----	448				
Db	57015	CTCTAAT---ATCTTGAGAAATGACCACATTTTCCCTAAGACGAGAGATGGATCTTCATGA	56959				
QY	449	-----SerMetArgAsnSerSerG	455				
Db	56958	TAATTTTCATGATGGTGCACTTCAAGAGGCTCATTTGTTGCTGCACAAATGTTTTCAGA	56899				
QY	455	uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyrSerThrAl	471				
Db	56898	ACAGCAGGAACCTGATAGCTGTTTCCAGGAATCATGGAATGCG---CATTTTCAACTGA	56842				
QY	471	aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysLeuLysLysG	489	TITLE			
Db	56841	TGAACCTCTCGCATTTGTGGAGAAAAGAAATCACTTAATTTGAATGCTTTCTCTGTAAG	56782	JOURNAL			
QY	489	nValGlnLysLeuValGlnGluHisGluLysLysMetGluLysLeuLeuAsnMetTyrAl	509	MEDLINE			
Db	56781	AGAAATCCGTATCTTTGATTAACACACAGTCTCAGATTTTCAGAGTTAGAGAAATACACACA	56722	PUBMED			
				REFERENCE			
				AUTHORS			
				TITLE			
				JOURNAL			

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-x@mail.nih.gov
 Tissue Procurement: Jeffrey B. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabhhu, Parvaneh Saeeidi, Jacqueline
 Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Scott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 86 Row: a Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES

source

Location/Qualifiers

1. .4190
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4983798"
 /tissue_type="Colon, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP Co24"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"
 <1. .4190
 /gene="Myo18a"
 /notes="synonyms: MyoP2D, MysP2D"
 /db_xref="LocusID:360013"
 /db_xref="MGI:2667185"

gene

CDS

<1. .3017
 /gene="Myo18a"
 /codon_start=3
 /product="Myo18a protein"
 /protein_id="AAH46638.1"
 /db_xref="GI:28436851"
 /db_xref="LocusID:360013"
 /db_xref="MGI:2667185"
 /translation="IKRSKMHFVHCFPVAEGWPGEPGRSARRVSSSELDLPDGP
 CEGALLQVLSLLRAQLRGSRLLDAMRYQVDPDMVFEFRFRDVLAPHLTKKHG
 RNVIIVDEKRAVEELLESLEKSSCCGLGSVFFPAGTLEBQORDQTSRHLTLF
 QAACGYLARQHPKKKIQLAIRCQKNIKNKGYKDPWPKLFTYVPLQVQLSE
 QRNRDESIQRLSKLEKVERNEIRUSSDRLETRIISLTSELDERTNTEGESASQ
 LDATERTLRITKEMKELQYDALKKQMEVMEVEARLIIRAAEINGEVDDDAAG
 EVRLKYRAVEDYFTKRLQLEEDKMEVEQSRRLERLGLQADSEORALQQ
 LKKKQRLTAELQDTKHLGEOVRNHEKKQRRFDSLSQAHESTOREKLOREKLO
 REKMLAEAFSLKQOMEKDLIDIAFTQKVVSLEAELODISSQESKEASLAKVKQ
 LRDLKAKDQBEELDEQAGSQMLEQAKRLMEMERWQTHSKEMESRDEVEAR
 QSQKQLKQMEVEEEDKQAKREKLESKLSLSDQNRQDFSEKRLKDLK
 RYKALLAQIIMDLHKNAPSKRETAIQLNQLSESEFTCAAAVKARKAMEVEDLH
 LQIDTAKAKTALIEQLRQEKNEIQNRLEEDQEDMNMELMKHKAQVAQASRDMAQ
 MNDLQAIIESNKKELOELQALQSOVEFLEQSMVDKSLVSROEAKTRLETRLEP
 EKTQVKLENLSRLKETMEKLTETEDORAAENKEQKBLQRLQRTKEEMSELA
 RKEASRKKHELMDELSLEAQLADLKLAFKIGDLQALAEIDENESDENEDLI
 NSEGDSDVSELEEDVDGVKSWLSKNKSGSKAPSDDGSLKSSSPTSHWKPLAPDPSDD
 EHPDVDSIRPFRSHSYLSDSDEAKLTETSA"
 645. .2654

misc_feature

/gene="Myo18a"
 /note="Myosin tail. Region: Myosin tail. The myosin
 molecule is a multi-subunit complex made up of two heavy
 chains and four light chains it is a fundamental

contractile protein found in all eukaryote cell types.
 This family consists of the coiled-coil myosin heavy chain
 tail region. The coiled-coil is composed of the tail from
 two molecules of myosin. These can then assemble into the
 macromolecular thick filament. The coiled-coil region
 provides the structural backbone the thick filament"
 /db_xref="CDD:pfam01576"

ORIGIN

Alignment Scores:
 Pred. No.: 0.00385 Length: 4190
 Score: 164.50 Matches: 137
 Percent Similarity: 38.10% Conservative: 119
 Best Local Similarity: 20.39% Mismatches: 227
 Query Match: 5.11% Indels: 190
 DB: 10 Gaps: 31

SEQ3_D1219_COPY_650_1270 (1-621) x BC046638 (1-4190)

QY 8 PheGlySerSerLeuAenProSerHisAspValLysHisValGluLysIleLysIleGlu 27
 Db 1284 TTCACAGTCAGCTTTCCAGGCCCCACGAGAGAGCGCGGAGAGCTG----- 1334
 QY 28 SerCysSerGluArgThrLysIlePheArgLeuTyrSerValCysThrValGluLys 47
 Db 1335 -----CAGAGCGAGAAACTCCACGCG-----GAGAAG 1361
 QY 48 AlaLeuIleLeuAlaArgGln---AenMetArgGlnAenLysAlaValGluAenLysAen 66
 Db 1362 GACATGCTCTGGCTGAGCTTTCAGCTTGAAGCAGCAATGAGGAGAAAGACTGGAC 1421
 QY 67 ---ArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyrLeuPheAspLysLeu 85
 Db 1422 ATTGCGGGGTTTACCAGGAAGGTGGTTCCTCTGGAGGCT-----GAGCTT 1466
 QY 86 AspHisPheHisSerSerGluThrProAspSerGly-----ValSerPheGlu 101
 Db 1467 CAGGACATTTCTTCTCAAGAGTCTAAGATGAGGCTTCTTGCCCAAGGTCAAGAGCAG 1526
 QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
 Db 1527 CTCGCGGACTTGGAGCCCAAGGTCAAGGACCAGGAGAGAGCTGGATGAGCAGGCTGG 1586
 QY 122 Glu-----GluAenGluValLysLeuCysLeuLeuGluAlaLysHisAlaGln 138
 Db 1587 AGTATACAGATGCTGGAGCAGCGCCAGCTCGCTCGAGATGAGATGGAGAGATGAGA 1646
 QY 139 GlyThrTyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGlu 158
 Db 1647 CAGACCCATTCCAAAGGAG-----ATGGAGAGTCCGGAGTGGATGAG 1682
 QY 159 GluSerProAenIlePheTrpSerLysLeuLeuGlyGlyLysAenProMetTrpLysTyr 178
 Db 1683 GAG----- 1685
 QY 179 ProSerAspThrProGlnArgAenArgLysArgValGlnTyrPheGluGlySerGluAla 198
 Db 1686 GTGGAAGAGCGCCGCGCAGTCATGTCAGAGAAGAGTTAAACACAGATGCAA---GTGCAGCTC 1742
 QY 199 SerProLysThrGlyAspGlyGlyAenAlaLysLysArgLysLys----- 213
 Db 1743 GAGGAGGAGTATGAAGACAAAGCAGAGGCGCTGCGGGAGAGAACCGGGAGCTGGAGCAAG 1802
 QY 214 -----AlaSerAspVal-----ThrAspProArgValThr 224
 Db 1803 CTCTCCACCTCAGTGCAGGATGACACCGCGGACTTTGAATCAGAGAGCGGCT-ACG 1861
 QY 225 AspProProValAspAspGluArgLysAlaSerGlyLys-----AspHisMetGly 242
 Db 1862 GAAAGACCT-----GAAGCGCACCAAGCGCTGCTGGCAGATGCCACATCATGCTGGA 1915
 QY 243 AlaLeu-----GluSerProLys-----GluSerProLys--- 248

ELLESLEKSSCGLSRVFPFAGTTLARLEBQRDRQTSRHLTLFOAACRGVILAROHF
 KRKKIQDLAIRCVQKIKKNGKVKMPWKLFTTVRPLIQVLSSEQIRNKDESIQOL
 RSKLEKVERKERNELRUSDRLETRISELTSERDERTGESASQLLDAETALRTEK
 ENKLEQTDALKQKQMEVMEYEARLIIRAAEINGEVDDDDAGGSRWLRKYRAVREV
 DFTKKQLQELDEKMEVQOQRRLRGLDQADSDSORALQQLKKCORLTABLO
 DTKLHLEGOVRNHELEKORRFRDSELSOAHEOTOREKLOREKLOREKMLLAEPFL
 KOOMEKQDLDIAGFTQKVVSLEAELODISQESKDEASLAKVKQLRDLBAKVQDOBE
 ELDEQAGSOMLEQAKLEMEMERMRQTHSKEMERDEBEAROSCKOKUKOMEVQ
 LBEYEDKQALREKRELESKLSLSDQVNRDFSEKRLKDLRKTALADAQIML
 DILKNAPSKRIEIAQKQLESEFTCAAVKARKAMEVEMEDLHLOIDIIAKAKTAL
 EQLSLRQEKNEIQNRLEBEDQDMNLMKKHKAQVAQSRDMAQNDLQAOIEENK
 EKQELQKLOALQSOVFELEQSMVDKSLVSROEAKIRELETRLEPEKTOVKRLENL
 RLKTEKLETERDQRAAENREKEQNKELQRLRDTKEMSELARKEAEARKKHEL
 EMDLEAANQSLQADLKLAFRIGDLQALAEDEMESENEDLINSBGSDSDVDSLE
 DRVDGVKSLSKNGKSKAPSDGSLKSSPSTSHWKPLAPDPSPDDEHDPVDSISPRF
 SHSYLSDSTAKLTETSA"

repeat_region

68..179

/note="KE repeat"

misc_feature

635..981

/gene="MysPDZ"

misc_feature

/note="PDZ domain"

1272..3358

/gene="MysPDZ"

misc_feature

/note="myosin head"

1544..1565

/gene="MysPDZ"

misc_feature

/bound_moiety="ATP/GTP binding site"

3602..3665

/gene="MysPDZ"

misc_feature

/note="IO domain"

3777..5865

/gene="MysPDZ"

polyA_signal

/note="coiled-coil"

7279..7284

/gene="MysPDZ"

polyA_site

7303

/gene="MysPDZ"

ORIGIN

Alignment Scores:

Pred. No.: 0.00823 Length: 7318
 Score: 164.50 Matches: 137
 Percent Similarity: 38.10% Conservative: 119
 Best Local Similarity: 20.39% Mismatches: 227
 Query Match: 5.11% Indels: 190
 DB: 10 Gaps: 31

SEQ3_D1219_COPY_650_1270 (1-621) x AB026497 (1-7318)

QY 8 PheGlySerSerLeuAsnProSerHisAspValLysHisValGluLysIleLysIleGlu 27
 Db 4426 TTCGACAGTGAGCTTCCAGGCGCCACGAGGAGCGCAGCGGAGAGCTG----- 4476
 QY 28 SerCysSerGluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLys 47
 Db 4477 -----CAGAGGAGAACTCCACGCG-----GAGAAG 4503
 QY 48 AlaLeuIleAlaAlaArgGln---AsnMetArgGlnAsnLysAlaValGluAsnLeuAsn 66
 Db 4504 GACATGCTCTGGCTGAGCTTTCAGCTTGAGAGCAATAGGAGGAGAAAGACTTGGAC 4563
 QY 67 ---ArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeu 95
 Db 4564 ATTGGCGGGTTTACCAGGAAGGTGGTTCTCTTGGAGGCT-----GAGCTT 4608
 QY 86 AspHisPheHisSerSerGluThrProAspSerGly-----ValSerPheGlu 101
 Db 4609 CAGGACATTTCTCTCAAGAGTCTAAGAGTGAAGCTTCTGTGCCCAAGGTCAAGAGCAG 4668
 QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
 Db 4669 CTCGGGACTTGGAGCCCAAGGTCAAGGACGAGGAGGAGGAGCTGGATGAGCAGCGTGGG 4728

QY 122 Glu-----GluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGln 138
 Db 4729 AGTATACAGATGCTGAGCAGCGCCAAAGCTCGCTCTGGAGATGGAGAGATGAGA 4788
 QY 139 GlyThrTyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGlu 158
 Db 4789 CAGACCCATTCGAGAG-----ATGGAGAGTCCGGATGAG 4824
 QY 159 SerSerProAsnIlePheTrpSerLysLeuGlyLysAsnProMetTrpLysTyr 178
 Db 4825 GAG----- 4827
 QY 179 ProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAla 198
 Db 4828 GTGGAAGAGCGCGCGGAGCTGTCAGAGAAGATTAACACAGATGGAA---GTGCAGCTC 4884
 QY 199 SerProLysThrGlyAspGlyGlyAsnAlaLysAlaLysArgLysLys----- 213
 Db 4885 GAGGAGGAGTATGAACAGCAGCAGAGAGCGCTGCGGGAGAAACGGGAGCTGGAGAGCAAG 4944
 QY 214 -----AlaSerAspAspVal-----ThrAspProArgValThr 224
 Db 4945 CTCTCCACGCTCAGTGACCGAGCTGAACACGCGGACTTTGAATCAGAGAAGCGCT-ACG 5003
 QY 225 AspProValAspAspAspGluArgLysAlaSerGlyLys-----AspHisMetGly 242
 Db 5004 GAAAGACCT-----GAAGCGCACCAAGCGCTGCTGGCAGATGCCAGATCATGTCTGA 5057
 QY 243 AlaLeu-----GluSerProLys--- 248
 Db 5058 CCACCTTAAGAACATATGCCCCGAGCAGAGGAGAGATGTCGCCAGCTGAAGAACCGCTGGA 5117
 QY 249 -----ValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAsp 266
 Db 5118 AGAGTCGGAATTCACCTGTGCAGCAGCTGTAAAGCCACGGAAGCGATGGAGTGGAGAT 5177
 QY 267 GlyAsnAspAla-PheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAs 286
 Db 5178 GGAAGACCTTCACCTGCAGATTTGATGACATCGCAAGCCAAAGCAGACGCTGGAGAGCA 5237
 QY 286 pMetLeuAlaSerGln---AspTrpGlyLysIlePro-----AspGluSerGlnAr 302
 Db 5238 ACTGAGTCGACTTCAGCGTGAGAAACAGAGATTTCAAACCGCGCTGAGAGAGATCAGGA 5297
 QY 302 gArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSe 322
 Db 5298 GGACATGATGAGCTGATGAAGAGCAGCAGCAGCTGTGCCAG-----GC 5345
 QY 322 rAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgI 342
 Db 5346 CTCCCGGACATGGCACAGATGAACGATCTCCAGGCTCAGATTGAAGAGTCCCAACAGGA 5405
 QY 342 eTyrGluGluProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla 362
 Db 5406 GAAGCAAGAGCTACAGGAGAGCTACAGCTCTGCAGAGCCAGGCTGGAGTTCCTCGAGCA 5465
 QY 362 aLeuLeuVal-----LysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerG 381
 Db 5466 GTCCATGTGGGACAGTCCCTCGTCAGCAGGAGGAGGAGGAGAGATCAGG----- 5514
 QY 381 uLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMe 401
 Db 5514 ----- 5514
 QY 401 tLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSe 421
 Db 5515 -----GAGCTGGAGACGACCTGGAGTTC----- 5538
 QY 421 rLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValAr 441
 Db 5538 ----- 5538
 QY 441 gAspGluLysIleAsnThrLysSerMetArgAsnSerSer-----GluAspG 457

```
Db      5539  -----GAAAAGAGCCCAAGTGAAGCGCTGGAGAACCTGGCGCAGTCGCTCAAAAGAAACCAT 5594
QY      457  uGluCyMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLy 477
Db      5595  GGAGAAAGCTGACTGAGGAACGGGACCAGCGCCGCGCTGAGAACCCGTGAG----- 5646
QY      477  sThrIleSerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHi 497
Db      5647  -----AAGGAGCAGAACCAAGAGGCTCCAGCGCACAGCTTCGTGACAC 5687
QY      497  eGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAsp-----LysLysGlnLy 514
Db      5688  CAAGGAGGAGATGAGCGAGCTTGCACAGGAAGGAGCGAGGCTAGCGCCGCAAGCATGA 5747
QY      514  sLeuGluThrSer---LysSerValGluAlaAla-----ValIleArgIleThrCysSe 531
Db      5748  ACTGGAGATGGACCTGGAGAGGCTGGAGCTGCTAACCAAGCCTGCAAGCCGACCTAAA 5807
QY      531  rArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluArgLysPh 551
Db      5808  GCTGGCGTTCAAAACGCATTGGGGACTTGCAAGCTGCCATTGAG----- 5850
QY      551  eAspGluIleLysSerGluLysAsnGluCysLysLysSerLeuGluGlnMethHisAspVa 571
Db      5851  -GATGAGATGGAAGTGAACGAGAACGAGGACCTCATCAACAGTGAGGGGGACTCAGATGT 5909
QY      571  lAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAl 591
Db      5910  GGACTCAGAG-----CTGGAGGCCGG-----GTCGACGGGGTCAAGTCTCTGGTT 5954
QY      591  aAlaLysLeuLys-----ValCysValProIleGlnSerGlyAsnAsnLysHisPheSe 609
Db      5955  GTCAAAGAACCAAGGGACCTTCTAAGGCACCTTCTGACGATGGCAGCTTGAAGAGTTCCAG 6014
QY      609  rGlySerSerAsnIleSerGlnAsnAlaProAsp 620
Db      6015  CCCAACCCAGCCACTGGAGCCCACTCGCCCTGAC 6048
```

Search completed: July 11, 2005, 19:44:15
Job time : 7902.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 09:34:25 ; Search time 865 Seconds
(without alignments)
4249.895 Million cell updates/sec

Title: SEQ3 D1219_COPY_650_1270

Perfect score: 3217
Sequence: 1 RADAFILFSSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q/cpn2 1/USPTO.spool/ROBINSON929/runat_11072005_102151_29159/app.query.fasta_1.2325
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=ROBINSON929 @CGN 1.1234 @runat_11072005_102151_29159 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:*

1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004as:.*
13: Geneseqn2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3213	99.9	6571	4	AAA89354 Arabidops
2	2977	92.5	10329	4	AAA89353 Arabidops
3	2975	92.5	10329	4	AAA89385 Arabidops
4	494.5	15.4	467	4	AAA89379 Brassica
5	469	14.6	515	4	AAA89381 Brassica

C	6	442	13.7	519	4	AAA89377
	7	164	5.1	5220	12	ADQ59166
	8	164	5.1	5837	12	ADL12346
	9	164	5.1	7764	5	ADL63336
	10	164	5.1	7764	10	AD54628
	11	164	5.1	7764	12	ADK60458
	12	164	5.1	7764	12	ADK60759
	13	164	5.1	7764	12	ADP73081
	14	160.5	5.0	4769	12	ADQ64522
	15	160	5.0	3045	13	ADS48467
	16	158	4.9	5858	2	AAQ10378
	17	158	4.9	5883	12	ADQ87119
	18	158	4.9	5883	13	ADQ84854
	19	158	4.9	5883	13	ACN40924
	20	158	4.9	6432	5	AA65903
	21	158	4.9	7274	13	ADN60282
	22	158	4.9	7396	6	ABT08505
	23	158	4.9	7396	12	ADO09868
	24	158	4.9	7516	4	AAK51987
	25	157	4.9	6354	10	ADU56352
	26	156	4.8	5175	2	AAH40200
	27	156	4.8	5896	4	AAH57404
	28	154	4.8	5857	6	ABL61968
	29	154	4.8	5857	10	ADB31363
	30	154	4.8	5857	13	ADP56352
	31	154	4.8	5943	4	AA622920
	32	154	4.8	6028	4	AAH72966
	33	154	4.8	6364	8	ACA04019
	34	153.5	4.8	6965	10	ADD29780
	35	153.5	4.8	7636	10	ACF58150
	36	153.5	4.8	7694	6	ABX04175
	37	153.5	4.8	7694	8	ABX10347
	38	153.5	4.8	7695	6	ABL68595
	39	153.5	4.8	7695	10	ACF58151
	40	153.5	4.8	7743	10	ACF58152
	41	153.5	4.8	7761	10	ACF58153
	42	153.5	4.8	8063	5	ADL63564
	43	149.5	4.6	5847	6	ABZ32324
	44	149	4.6	5641	6	ABZ32359
	45	147	4.6	7985	4	ABL08391

ALIGNMENTS

RESULT 1
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
XX AC AAA89354;
XX DT 23-APR-2001 (first entry)
XX DE Arabidopsis thaliana silencing gene cDNA.
XX KW Gene silencing; silencing gene; MOM; ss.
XX OS Arabidopsis thaliana.
XX FH Key Location/Qualifiers
XX FT CDS 310..6315
XX FT /*tag= a
XX PN WO200100801-A2.
XX PD 04-JAN-2001.
XX PF 21-JUN-2000; 2000WO-EP005761.
XX PR 23-JUN-1999; 99GB-00014623.
XX PA (NOVS) NOVARTIS AG.
XX PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;

XX WPI; 2001-137952/14.

DR P-PSDB; AAB20062.

XX PT Novel gene encoding a protein that controls gene silencing, in particular

PT silencing of plant genes.

XX Claim 6; Page 22-32; 48pp; English.

XX The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with AtPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hydrogromycin resistance gene
CC (hvg). Mutation of a silencing modified locus results in release of
CC silencing of the hvg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression

XX SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.19e-298 Length: 6571
Score: 3213.00 Matches: 620
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.84% Mismatches: 0
Query Match: 99.88% Indels: 0
DB: 4 Gaps: 0

SRQ3_D1219_COPY_650_1270 (1-621) x AAA89354 (1-6571)

QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyHis	20
DB	2257	CGAGCTGATCGCTTTATCTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	2316
QY	21	ValGluLysIleLysIleGluSerCysSerGluAgtThrLysIlePheArgLeuTyrSer	40
DB	2317	GTAGAGAAGATAAATCATGATGTTCTGAAAGAACTTAGATATTCGATGTACTCA	2376
QY	41	ValCysThrValGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys	60
DB	2377	GTATGTACAGTTGAAGAAAAGCCCTGATCTCGCTAGGCAGCAAAATATGCGCAAAATAAA	2436
QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
DB	2437	GCTGTAGAGAACCTAAACCGCTCTCTCACGCACGCACTGCTCATGTGGGGGGCGTCATAC	2496
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
DB	2497	TTATTTGATTAACCTGGATCATTTTACAGCAGTGAACCTCCAGATTCCAGAGTTTCATTT	2556
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
DB	2557	GAACAATCTATTATGACGGCGTGATTCATGAATTCCTGTCATCTTTCTTCCAAAGGT	2616
QY	121	GlyGluGluAsnGluVallyLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr	140
DB	2617	GGAGAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAACT	2676
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
DB	2677	TACAGCAGTGATTTCTACTCTATTGTGTGAAGCACTATTAAAGTTGTACAGATGAAGAGT	2736

QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer	180
DB	2737	CCAAATATATTTTGGTCAAGAGCTGTTGGGGGAAAAAATCCCTATGTGGAAATACCCCTTCA	2796
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
DB	2797	GATACTCCCCAAAGGAATCGAANAACGAGTTTCAGTATTTTGGAGGTTCGAGCGAGTCCC	2856
QY	201	LysThrGlyAspGlyLysAlaLysLysArgLysLysAlaSerAspValThrAsp	220
DB	2857	AAAACTGGCGATGTTGGAAATGCAAGAAGCGAAAGAGGCTTCTGATGATCACTGAT	2916
QY	221	ProArgValThrAspProValAspAspGluArgLysAlaSerClyLysAspHis	240
DB	2917	CCCCGGGTCACTGATCCGCCAGTAGATGATGAAGAAAGCCCTCTGGGAAGATCAC	2976
QY	241	MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly	260
DB	2977	ATGGGGGCTTTGGAGTCAACAAAAGTCATAACATCCAGTCATCATGTAAATCTTCTGAT	3036
QY	261	ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle	280
DB	3037	ACAGATGGTACATTCGATGGAAATGATGCTTTTGGCTTGTAATCTATGGGAGCCATATC	3096
QY	281	SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer	300
DB	3097	TCGTGAATCCCAAGAGATATGTTAGTATCGAAGATGGGGGAAAAATACCCGATGAATCA	3156
QY	301	GlnArgArgLeuHisThrValLeuLysProLysMetAlaLeuLysCysGlnValLeuHis	320
DB	3157	CAGAGAGGCTCCACACTGTTTAAAGCCGAAGATGGCAAACTTTGCCAAGTTTTCAT	3216
QY	321	LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis	340
DB	3217	CTTTCAGATGCTTGCAAGCATGCTCGGAAATTTTCTCGAATATGTTATTGAAATATCAC	3276
QY	341	ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle	360
DB	3277	CGAATCTACGAAGAGCCGCACTACTTTTCAGGCATTTCCAGATAGCCCTGATGTGGATT	3336
QY	361	AlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer	380
DB	3337	CGAGCTTGTGTTGGTAAGCAAAATCTTAGCCCAAGAAATCTCTGGTCGCGCAAAATCTC	3396
QY	381	GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys	400
DB	3397	GAATTAGCTTTCAATGCTCTAGAGTAGAGGTGGATTATATTTTCGATATTGTCCTGC	3456
QY	401	MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn	420
DB	3457	ATGAAGAGTCTGTTCTCGAGCATACCAAGGTTTTCAGATTGCTGATGCTTTCGTTACTAAT	3516
QY	421	SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal	440
DB	3517	TCTAAACAGTCACTGTTGTAGCAAAAATGATGTAATGAAGTCTCTCAGGGGCTACAGTG	3576
QY	441	ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluCysMet	460
DB	3577	CGTACGAAAAGATTAAATACGAAGTCGATCGCAATAGCTCAGAGGATGAAGAGTGCATG	3636
QY	461	ThrGluLysArgCysSerHisThrSerThrAlaThrArgAspIleGluLysThrIleSer	480
DB	3637	ACTGAGAAGAGATGTAGCCATTATAGCACAGCAAGAGATATCGAAAAGACTATTAGT	3696
QY	481	GlyIleLysLysLysLysLysGlnValGlnLysLeuValGlnGlnHisGluGluLys	500
DB	3697	GGCATAAAAAGAAATACAGAGCAAGTGCAAAAGCTTGTACAGAGCATGAGGAAAG	3756
QY	501	LysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer	520
DB	3757	AAAAATGGAGCTGTTAAATATGTTATGCAGCAAGAGAGCAAACTTGAAACTAGTAAAGT	3816
QY	521	ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu	540

Db 5550 GGCTGATCAAGACATGTCAAAAGTTTATCTGTTCGCTGTGGTATAGAAAATAATACAG 5609
 QY 322 ----- 322
 Db 5610 TGTCTGATGCTATTTTAAAGGTAGTCTTATGCTTTCACATATTCGCTAATAGATGTTTCCG 5669
 QY 322 ----- 322
 Db 5670 CTGTGCTGCCATATCTTCTGTGATTCACGGTGTCCGTCTATCAAAAATTGTACTAA 5729
 QY 323 ----- AspAlaCysThr 326
 Db 5730 AAGTATTTTGCATGCTGATTCGTTAAACAGATTATTTTGTTCAGGATGCTGCACA 5789
 QY 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
 Db 5790 AGCATGTCGGAATTTTCTCGAATATGTTATTGMAAATCACCGAATCTACGAAGAGCCA 5849
 QY 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
 Db 5850 GCCACTACTTTTTCAGGCAATTCAGATAGCCCT-GGTATGACAGCATTTTACTTTGTATAATT 5908
 QY 357 ----- 357
 Db 5909 TATGCAATTGTTCTTCATCATCTGCTTTGTTTGTAGAAATGCTCTCAGAGGACGACTCC 5968
 QY 357 ----- 357
 Db 5969 TTTAGTTTTAACTTTCGAATCATAGGATTCAAATATCCATTAAGTGCCTTTTGATCGCTG 6028
 QY 358 ----- SerTyr 359
 Db 6029 CATAATATATGAATAGTTGACATCTGAATACGTTGTTAATAATGCAATTTTCAGAGTTGG 6088
 QY 360 IleAlaAlaLeuLeuValIleGlnIleLeuSerHisGluSerLeuValArgAlaAsn 379
 Db 6089 ATTGCAAGCTTGTGGTAAAGCAAAATCTTTAGCCACAAAGAAATCTCTGGTCCGTGCAAT 6148
 QY 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
 Db 6149 TCTGAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGGATTATATTATTCGATTATGTCC 6208
 QY 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
 Db 6209 TGCATGAAGAGTCTGTTCTCTGGAGCATACAAAGTTTGCAGTTTCGATTCTTTGGTACT 6268
 QY 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr 439
 Db 6269 AATCTTAAACAGTCAGTGGTTAGCACAAAACCTAGTAAATGAAAGTCTCTCAGGGGCTACA 6328
 QY 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
 Db 6329 GTGCGTGACGAAAGATTATACGAAGTCATCGGAAATAGCTCAGAGGATGAAGAGTGC 6388
 QY 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
 Db 6389 ATGACTCAGAAGAGATGTAGCCATTATAGCACAGCAACAGAGATATCGAAAAGACTATT 6448
 QY 480 SerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
 Db 6449 AGTGGCCATAAAAAGAAATCAAGAACAAAGTGCAAAAGCTTGTACAAGAGCATGAGGAA 6508
 QY 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
 Db 6509 AAGAAATGAGCTGTTAATATATGATGCAGACAGAGAGCAACTTGAACCTAGTAAA 6568
 QY 520 SerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
 Db 6569 AGTGTGAAGCAGCAGTAATTCTGATTATCTGTTTACCGGACCACTCAAGTGGGTGAT 6628
 QY 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn 559
 Db 6629 CTCAACTGCTGGATCATATATTATGAAGAAAGTTTGTATGAAATCAAAAGTGAGAAAAT 6688

QY 560 GluCysLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGluAsp 579
 Db 6689 GAATGCTCAAAAGTCTGGAGCAATGCACAGGTTGCAAAAGAAAGTTGGCTGAGGAT 6748
 QY 580 GluAlaCysTrpIleLeuAsnArgIleLysSerTrpAlaAlaLysLeuValCysValPro 599
 Db 6749 GAAGCTGTGTGGATTAAATCGATAAAGAGCTGGCAGCTAAATTTAAAGTTTGTGTCCC 6808
 QY 600 IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro 619
 Db 6809 ATTCAAAGTGGCAATTAACAAGCATTTTAGTGGTTTCATCAACATTTTCCCAAAATGCTCCT 6868
 QY 620 AspVal 621
 Db 6869 GATGTA 6874
 RESULT 3
 AA889385
 ID AAA89385 standard; DNA; 10329 BP.
 XX AC AAA89385;
 XX 04-NOV-2004 (revised)
 DT 23-APR-2001 (first entry)
 XX
 DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
 XX
 KW Gene silencing; silencing gene; MOM; da.
 XX Arabidopsis thaliana.
 OS
 FH Key Location/Qualifiers
 FT exon 1. .1008
 FT /tag= a
 FT /number= 1
 FT intron 1009. .1295
 FT /tag= b
 FT /number= 1
 FT exon 1296. .2550
 FT /tag= c
 FT /number= 2
 FT CDS 1301. .9808
 FT /tag= d
 FT /product= "Columbia protein"
 FT /note= "This sequence contains 15 introns."
 FT intron 2551. .2673
 FT /tag= e
 FT /number= 2
 FT exon 2674. .2752
 FT /tag= f
 FT /number= 3
 FT intron 2753. .2867
 FT /tag= g
 FT /number= 3
 FT exon 2868. .3113
 FT /tag= h
 FT /number= 4
 FT intron 3114. .3506
 FT /tag= i
 FT /number= 4
 FT exon 3507. .3680
 FT /tag= j
 FT /number= 5
 FT intron 3681. .3973
 FT /tag= k
 FT /number= 5
 FT exon 3974. .4895
 FT /tag= l
 FT /number= 6
 FT allele 4338
 FT /tag= m
 FT /note= "ecotype Zurich"

Db	4532	GGAGAGAAATGAAGCTGAGCTGCTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAC	4591	Db	5610	TGCTCATGCTATTTTAAGGTAGTCTTATGTCTTCACATATGGCTAATAGATGTTTCCG	5669
Qy	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160	Qy	322	-----	322
Db	4592	TACAGCAGTGATCTACTCTATTGTTGGTGAAGACATATTAAAGTTGTCAAGATGAAGAGT	4651	Db	5670	CTGTCGTGCCATATACTTCTGTGATTATCACGGTGTCCGTCTATCAAAATTTGTACTAA	5729
Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpSer	180	Qy	323	-----	326
Db	4652	CCAAATATATTTGGTCAAGCTGTTGGGGGAAAAATCCTATGTGGAATACCTTCA	4711	Db	5730	AAGTATTTTGCATGTGTGATTGGTTAAACAGATTATTTTGTTCAGGATGCTTGACA	5789
Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200	Qy	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro	346
Db	4712	GATACTCCCCAAAGGAATCGAAACAGATTACGATATTTTGGAGGTTCTGAAGCGAGTCCC	4771	Db	5790	AGCATCGTCGGAATTTTCTCGAATATGTTATTGAAATACCCGAATCTACGAAGAGCCA	5849
Qy	201	LysThrGlyAspGlyGlyAsnAlaLysLysArgLysAlaSerAspValThrAsp	220	Qy	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu	357
Db	4772	AAAACTGGCGATGTTGGAATGCAAGAGCGGAAGAGGCTTCTGATGATGATCATGAT	4831	Db	5850	GCCACTACTTTTTCAGGCATTCAGATAGCCCT-GGTATGACAGACATTACTTTTGATAATT	5908
Qy	221	ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis	240	Qy	357	-----	357
Db	4832	CCCCGGTCACTGATCCGCCAGTAGATGATGATGAAGAAAGGCCCTCTGGGAAGGATCAC	4891	Db	5909	TATGCAATGTTTCCCTTCATCATCTGCTTTGTTTGAATGTCTCAGAAAGCAGCAGCTCC	5968
Qy	241	Met	241	Qy	357	-----	357
Db	4892	ATGGGTAAAAATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA	4951	Db	5969	TTTAGTTTAACTTTTCCAAATCATAGGATTCAAATATCCATTAATCTGCCTTTGATCGCTG	6028
Qy	242	-----	253	Qy	358	-----	359
Db	4952	CTTTGCTGACTATCTTTCCTACAGGGGCT-TTGGAGTCCACAAAGTCATAACACTCCAG	5010	Db	6029	CATAATATATGAATAGTAGTGCATACACTGTAATACGTTGTTAATATGATTTTCAGAGTTGG	6088
Qy	254	SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273	Qy	360	IleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn	379
Db	5011	TCATCATGTAATCTTCTGTTACAGATGTTACATTTGGATGTAATGATGCTTTTGGCTTG	5070	Db	6089	ATTTCAGACCTTGTGTTAAAGCAAAATCTTAGCCACAAAGAAATCTCTGTCCTCGCAAA	6148
Qy	274	TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293	Qy	380	SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer	399
Db	5071	TATTTCTATGGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGTAGTCAAGATTGG	5130	Db	6149	TCTGAATTAGCTTTCAATGCTCTAGAGTAGAGTGGATTATATTTATTCGATATTGTTC	6208
Qy	294	GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla	313	Qy	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419
Db	5131	GGGAAATACCGGATGAATCACAGAGAGGCTCCACATGTTTTTAAAGCCGAAGATGGCA	5190	Db	6209	TGCATGAAGAGTCTGTCTCTGGAGCATACACAAGGTTTGCAGTTCCGATTGCTTGTACT	6268
Qy	314	LysLeuCysGlnValLeuHisLeuSer	322	Qy	420	AsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr	439
Db	5191	AAACTTTGCCAAGTTTTCATCTTC-AGTAAAGTGGCTTTTTCACCTCCACAACTTATT	5249	Db	6269	AATTTCTAAACAGTCACTGTTAGTACCAAACTAGTAAATCAAAGTCTCTCAGGGGTACA	6328
Qy	322	-----	322	Qy	440	ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys	459
Db	5250	TTAGCCTTGCAATATGCTTATATATAGCTGATTGCAACTGTAGTTGTTTACCTGATTTTCTG	5309	Db	6329	GTGCGTGACGAAAGATTAAATACGAAGTCGATCGGAAATAGCTCAGAGGATGAAGAGTGC	6388
Qy	322	-----	322	Qy	460	MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle	479
Db	5310	TTACAGCCAAATGTGAGAGTTTTTATCTTCAACTATATCCATCCGTTTAAGCATATTTTA	5369	Db	6389	ATGACTGAGAGAGATGTAGCCATTATAGCACACGCAACAGAGATATCGAAAGACTATT	6448
Qy	322	-----	322	Qy	480	SerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu	499
Db	5370	TTTCTTATATCTGGCTTCGTTACCAATGCACTGTTAAATGAGCAACTGTCGACAAAC	5429	Db	6449	AGTGGCATAAAGAAATACAAAGCAAGTCAAGTCAAGTCAAGAGCATGAGGAA	6508
Qy	322	-----	322	Qy	500	LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys	519
Db	5430	AGTAGGTAGTTATGTGCTCATGTCAATTCATTGTTTATTGAAGCAAAAGAAATTTCTGTCT	5489	Db	6509	AGAAATATGGAGCTGTAAATATGTTATGATGACAGCAAGAGCAGAACTTGAACACTAGTAA	6568
Qy	322	-----	322	Qy	520	SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539
Db	5490	ACTTTACATGATCCATCTGTGGGAGTATATACTATATATAACCTTAGGCCCTTTGTACCT	5549	Db	6569	AGTGTGAAGCAGCAGTAAATTCGTATTACTGTTTACGGACCGACTACTCAAGTGGGTGAT	6628
Qy	322	-----	322	Qy	540	LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn	559
Db	5550	GGCTGATCAAGACATGTCAAAAGTTTATCTGTTGCTGTTGTTGATAGAACTAATACAG	5609	Db	6629	CTCAAACTGCTGATCATTAATATGAAGAAAGTTTGTATGAATCAAAGTGAAGAAAAAT	6688
Qy	322	-----	322	Qy	560	GluCysLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGluAsp	579
Db				Db	6689	GAATGCTCAAAAGTCTGGAGCAAAATGSCAGATGTTTGCAAAGAAAGTGTGGCTGAGGAT	6748

```

QY 580 GluAlaCysTrpIleAsnArgIleYsSerrTrpAlaAlaLysLeuLysValCysValPro 599
Db |||||
6749 GAAGCTGTGGATTAAATCGGATAAAGAGCTGGCAGCTAAATTAAGTTTGTGTTCC 6808
QY 600 IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro 619
Db |||||
6809 ATTCAAAGTGGCAATAACAAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCT 6868
QY 620 AspVal 621
Db |||||
6869 GAATGA 6874

RESULT 4
ID AAA89379 standard; DNA; 467 BP.
XX
AC AAA89379;
XX
DT 11-SEP-2003 (revised)
DT 23-APR-2001 (first entry)
XX
DE Brassica oleracea silencing gene homologous DNA seq1-43.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Brassica oleracea; acephala.
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
XX
DR Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Example 6; Page 45; 48pp; English.
XX
CC This is the sequence of Seq1-43, a genomic DNA sequence of Brassica
CC oleracea acephala that shows homology to the newly identified silencer
CC gene (see AAA89379) of Arabidopsis thaliana. Seq1-27 is a partial
CC sequence of clone 1, which was isolated from a genomic library screened
CC with MOM cDNA. Partial sequences of clone 1 (see also AAA89377-80) showed
CC similarity to different regions of the MOM gene (80-86% at DNA level and
CC 62-80% at amino acid level) which encode the N-terminal, Arpase, and C-
CC terminal parts of the MOM protein (see AAB20062). Gene silencing is
CC useful as a molecular tool for regulating gene expression. (Updated on 11
CC -SEP-2003 to standardise OS field)
XX
SQ Sequence 467 BP; 145 A; 87 C; 108 G; 127 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,69e-38 Length: 467
Score: 494.50 Matches: 97
Percent Similarity: 84.09% Conservative: 14
Best Local Similarity: 73.48% Mismatches: 12
Query Match: 15.37% Indels: 9
DB: Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x AAA89379 (1-467)

QY 111 GluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeuCysLeu 130
Db |||||

```

```

Db 1 GAATTCTCGTCATATCTTCTTCCGATGTTGGAGAGAAATGAAGGCAAGCTGTGCTA 60
QY 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTySerSerAspSerThrLeuPheGlyGlu 150
Db |||||
61 CTTTGGGAAGCAAGCATGCTCAGGGAAGTTACAGCACTGATGCTACTTATTGGTGAA 120
QY 151 AspHisIleLeuLeuSerAspGluSerProAsnIlePheTrpSerLysLeuLeuGly 170
Db |||||
121 GAACATGTCAAAGTATTCAGAT---GAAAGTCCAAATATGTTTGGTCAAAGCTGTGAGT 177
QY 171 GlyLysAsnProMetTrpLysTyProSerAspThrProGlnArgAsnArgLysArgVal 190
Db |||||
178 GGAAAGAAACCCCTATGTGGAATACTGTTCCGGATATCTCTCAAAGGAGTCAAAAGAGTA 237
QY 191 GlnTyPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLys 210
Db |||||
238 CGGCATCTTCAGGGCTATGAGGAGACTACCAAAGTTGGCAATGGCGGAAACTTAAAGAAG 297
QY 211 ArgLysLysAlaSerAspValThrAspProArgValThrAspProProValAspAsp 230
Db |||||
298 AAAAAGAGGGCTTCAGATGATGTCACA-----GTAGATAAC 333
QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGly 242
Db |||||
334 GCTGAGAGAAAGCCCTCTGGAAAGGATCACATGGGT 369

RESULT 5
ID AAA89381 standard; DNA; 515 BP.
XX
AC AAA89381;
XX
DT 11-SEP-2003 (revised)
DT 23-APR-2001 (first entry)
XX
DE Brassica oleracea silencing gene homologous DNA seq2-33.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Brassica oleracea; acephala.
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
XX
DR Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Example 6; Page 46; 48pp; English.
XX
CC This is the sequence of Seq2-33, a genomic DNA sequence of Brassica
CC oleracea acephala that shows homology to the newly identified silencer
CC gene (see AAA89381) of Arabidopsis thaliana. Seq2-33 is a partial
CC sequence of clone 2, which was isolated from a genomic library screened
CC with MOM cDNA. Partial sequences of clone 2 (see also AAA89382-84) showed
CC similarity to different regions of the MOM gene (64-76% at DNA level and
CC 55-64% at amino acid level) which encode the Arpase, putative
CC transmembrane and C-terminal parts of the MOM protein (see AAB20062).
CC Gene silencing is useful as a molecular tool for regulating gene
CC expression. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 515 BP; 181 A; 74 C; 136 G; 124 T; 0 U; 0 Other;

```

Alignment Scores:
 Pred. No.: 1.23e-35 Length: 515
 Score: 469.00 Matches: 99
 Percent Similarity: 60.87% Conservative: 27
 Best Local Similarity: 47.83% Mismatches: 39
 Query Match: 14.58% Indels: 42
 DB: 4 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x AAA89381 (1-515)

QY 387 SerArgValGluValAspTyrIleThrSerIleLeuSerCysMetLysSerLeuPheLeu 406
 DB 1 TCTAGAGAGAGGTGGATTATATATCTTTCTGTACTGCATGAAGATCTATTGCTG 60
 QY 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
 DB 61 GGGCGCACACAGGTTCCAA----- 81
 QY 427 SerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn 446
 DB 81 ----- 81
 QY 447 ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSer 466
 DB 82 -----GAAAGGGGTGAAGATGCGTGCAGAAAGAGGTAGC 120
 QY 467 HisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTyr 486
 DB 121 CATATAGCTCAGTACCAAGAGTGTGAAAGACTATTAGCCACATCAAAAGAAATGC 180
 QY 487 LysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsn 506
 DB 181 AGTAAGAGCCTGCATTAAGCTTTGTACAAACCTCGAGGAGAAAGATGGACCTGTGAAAT 240
 QY 507 MetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaValIle 526
 DB 241 AGGAATGCTGTCAAGAGCAGGAACTTCAGATTGTAAAGGTGGAAGCATCATTTATT 300
 QY 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
 DB 301 CGTGCACCTATTTCAGTATATAATCTCAGAGCTTACATGATGCTCTCCAACGGCTGGAA 360
 QY 545 HisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSer 564
 DB 361 TGTACTTTTGAAGAAAGTTTGATGATCTCAAGAGGAGGTGGATGAATGCTTGAAGT 420
 QY 565 LeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
 DB 421 TTAGAGCAATAAACCGAGGCTGGAAAGAGAGTGTGGCTGAAGATGAAGCCTGTGGATT 480
 QY 585 AsnArgIleLysSerTrpAla 591
 DB 481 AGTCGGATAGAGAAATGGGCA 501

RESULT 6

AAA89377/c
 ID AAA89377 standard; DNA; 519 BP.
 XX
 AC AAA89377;
 XX
 DT 11-SEP-2003 (revised)
 DT 23-APR-2001 (first entry)
 XX
 DE Brassica oleracea silencing gene homologous DNA seq1-23.
 XX
 KW Gene silencing; silencing gene; MOM; ds.
 XX
 OS Brassica oleracea; acephala.
 XX
 FN WO200100801-A2.
 XX
 PD 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.
 XX 23-JUN-1999; 99GB-00014623.
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 XX WPI; 2001-137952/14.
 XX Novel gene encoding a protein that controls gene silencing, in particular
 PT silencing of plant genes.
 XX Example 6; Page 44; 48pp; English.
 XX This is the sequence of Seq1-23, a genomic DNA sequence of Brassica
 CC oleracea acephala that shows homology to the newly identified silencer
 CC gene (see AAA89353) of Arabidopsis thaliana. Seq1-23 is a partial
 CC sequence of clone 1, which was isolated from a genomic library screened
 CC with MOM cDNA. Partial sequences of clone 1 (see also AAA89378-80) showed
 CC similarity to different regions of the MOM gene (80-86% at DNA level and
 CC 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-
 CC terminal parts of the MOM protein (see AAB20062). Gene silencing is
 CC useful as a molecular tool for regulating gene expression. (Updated on 11
 CC -SEP-2003 to standardise OS field)
 XX SQ Sequence 519 BP; 150 A; 117 C; 101 G; 149 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 4.98e-33 Length: 519
 Score: 442.00 Matches: 89
 Percent Similarity: 84.82% Conservative: 6
 Best Local Similarity: 79.46% Mismatches: 13
 Query Match: 13.74% Indels: 4
 DB: 4 Gaps: 1

SEQ3_D1219_COPY_650_1270 (1-621) x AAA89377 (1-519)

QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
 DB 324 CGCGCTGAGTCTTTTATCTTTTGGAGCAGCCTCAATCCATCGCATGACGTTAAGCTC 265
 QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
 DB 264 TTAGAGAAATAAAGTGTGCTGCTTCTGAAGAACTAAGATTTCCGGTTGACTCA 205
 QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLys 60
 DB 204 GTGTTTACAGTTGAAGAAAAACCCCTGATTCTTGCT-----AGGCAAAATAAG 157
 QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 DB 156 CCTCTAGAGAACCTAAACCGTCTCTCACACATGCATTCATGTGGGGGATCATAT 97
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 96 TTTATTTGATAAGCTGGACACTTCCATGGAGTGGAAACCCAGGTTCCAGAGTTCCATT 37
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
 DB 36 GAACAGTCTATTAAAGGATGCCGTAAANNCAGGAATTC 1

RESULT 7

ADQ59166
 -ID ADQ59166 standard; cDNA; 5220 BP.
 XX
 AC ADQ59166;
 XX
 DT 09-SEP-2004 (first entry)
 XX
 DE MSI-H carcinoma cDNA sequence SEQ ID NO:3.

Db 2760 GCGGACACAGTCGTCTATCTTTGACTCTGACTGGAACCCACAGAAATGACTTGGAGGCACAA 2819
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2820 GCCCGAGCCATAGATTGGTCAGAGAGACAGGTAAATATTACCGCTTAGTTCACAAAG 2879
Qy 42 CysThrValGluGluLysAla 49
Db 2880 GGGACTGTGGAGGAGAGATCATAGAACGGGCCMAAAGAGAGATGTTATTAGATCATCTG 2939
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn 64
Db 2940 GTGATTACGCGCATGACACCACTGCGCGGACGATCTCGAAACAACTCAGGAAGTCC 2999
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3000 AACTCAATCCCTTTTAAAGAGAGCTGACACGCTATTTTGAATTTGGAGCAGGAT 3059
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3060 CTCTTCAAGAACTGGAA -----GGGGAGGAATCA 3089
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3090 GAACCTCAGGAATGGAT -----ATAGATGAATTTTTCGGTTGGCTGAA 3134
Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3135 ACAGAGAGAGATGAAGTGTACACAGTGCACAGATGAATCTTA ---TCACAGTTTAG 3191
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3192 GTTGCCAACTTTGCCAAATGGAAGATGAAGAAGACTA -----GAAGAGCGT 3239
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 3240 CCTCAAGAGCTGGGATGAGATCAT ----- 3266
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
Db 3267 -----CCAGAGGACAAAGGAAAGATAGAGGAGAGAGCGGCAGAGGAGTAGAA 3320
Qy 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3321 GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAAAGGCTCAGACAAATGACAGT 3380
Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
Db 3381 GACTCTGACACTGAGTCTAAGAGGCGAGCCCGCAGAGATCCTCTGCTTCTGAGAGTGAACG 3440
Qy 224 ThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 3441 GAAGAC -----TCTGATGATGACAGAAG ----- 3464
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 3465 -----CCAAAGCGCAGAGGCGCTCCGAGAGTGTGCGAAGGACCTCGTGGAGGGA 3515
Qy 264 ThrLeuAspGlyAsn -----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 3516 TTTACTGATGCGAGATCCGAAGGTTTCATCAAGGCTTATAAGAGTTT ----- 3563
Qy 279 HisIleSerGlyIleProGluAspMetLeu -----AlaSerGlnAspTrpGlyLysIle 296
Db 3564 -----GGTCTCCCTCTTGAAACGCTGGAGTGCATGACAGCTGAT ---GCTGAGCTG 3611
Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 3612 GTAGATAAGTCGGTGCAGATCTG -----AAGCGCTGGGT 3647
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 3648 GAACGTATCCAC -----AACAGCTGTGTGTGCAATGCAAGGAAATACGAAGAGCAGCTG 3701

Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 3702 AAAGAAATATCCAGCAGGAAAGGACCAGGAAAGG ---AGAGTCCAACAAATCAAG 3758
Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 3759 ATATCCGAGTT ---CAGTTAATGTGAATCCATTATCCAAACATGAAGAGAGTTT --- 3812
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal -----GluValAspTyrIle 394
Db 3813 -----GAGATGCTGCATAAATCTATCCCTGTGGACCTCGAAGAAAAA 3860
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 3861 TACTGCTTAACCTGTCGTCAAGCTGCACATTTTGTAGTAGTGGGGGTGGAAGAT 3920
Qy 415 AspCys -----PheGlyThrAsnSer -----Lys 422
Db 3921 GATTCCTGCTGTTGCTGGGATTTATGAACATGCTATGGAACACTGGGAGTTAATAA 3980
Qy 422 sGlnSerValValSerThrLysLeuValAsn ----- 432
Db 3981 ACAG -----ACCAGAGCTTAAATTAATCTGACAAATTTCTGCCGTGGAGACA 4028
Qy 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn ---ThrLysSerMet 450
Db 4029 GATAAAAGCCTCAGGGAGCAGCTACAGACCCGAGCGATTACTTGTGAAGCTGCTC 4088
Qy 451 ArgAsnSerSerGlu -----AspGluGluCysMetThrGluLys 463
Db 4089 AGAAGAGGCTCTGGAAGAAGAGGGGCTGTGACAGGTGGGGAAGAGGCCAAATTAAGAAG 4148
Qy 464 Arg ----- 464
Db 4149 CGGAAGCCTCGGTAAGAGGAAACAAAGTCCAGGCTCAAGAGGAGCATGGAATT 4208
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrLeuSer 480
Db 4209 GAGCTTTCATCTCCTAGGCATTTCAGATAATCCATCAGAAGAGGGAGAGTGAAGATGAT 4268
Qy 481 GlyIleLysLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4269 GCGTTGGAAAAAGTCCAATGAAAAAACAAGAG ---AAGAAAGAGAACCAAGGAGAAC 4325
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys ----- 511
Db 4326 AAGGAGAAACAAATGAGTTCTAGGAAGACAAAGAAAGGGGACACAGGAAGAAAGATCA 4385
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla ----- 523
Db 4386 AAAGATAAGAAAGAGAGAGCCTAAAAAGTGTGTATGTCGCAATCTTCGAGTAAATCAAAGCGA 4445
Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp ----- 539
Db 4446 TCTCAGGCTCTGTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGAGATGAG 4505
Qy 539 ----- 539
Db 4506 GATGATGATCTGGACCGAGAGACATTCAGCATATGTNAGGAGAGGATGAGGCCGTGAAA 4565
Qy 540 -----LeuLysLeuLeuAspHis -----AsnTyrGluArgLysPheAsp 552
Db 4566 AAGGCACCTGAAACAGCTGCACAAACCTGCACAAAGGGGCTCAACGTGCAGCAAGCAGCTGGA 4625
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer --- 564
Db 4626 CACACCCGGAACCTGCTGTGTAATAATCGGAGACCGGATAGCCGAGTTCCTTAAAGCCTAC 4685
Qy 564 ----- 564
Db 4686 TCAGATCAGGAGCACATCAAACTCTGAGGAGGAACTATGATTTTTTTTCCAAAGTTT 4745

Qy	565	-----LeuGluGlnMethHisaspValalaLysLysLysLeuAla	577
			::::
Db	4746	ACGAATTTGATGTCGAAAACCTCATAAAGTTATACAAGATGGCTATAAGAAAAGTGCT	4805
Qy	578	GluaspGluAlaCysTrpIleAsnArgIleLysSerTrpalaIalLysLeuLysValCys	597
		::::::	
Db	4806	CAAGAAGAAGAG-----GAGCAAAAAGACAAAGAACGACGAC	4838
Qy	598	ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn	613
			::::
Db	4839	GTG-----ACTGGGGGTAAAGAAACCATTTGCTCAGAGGCTCAGGCTCCAGCGGG	4889
Qy	614	-----IleSerGlnAsn	617
			::::
Db	4890	GACTCTCTGATATCTCAGTGCC	4910

```

RESULT 9
ADL6336
ID ADL63336 standard; DNA; 7764 BP.
XX
XX
XX ADL63336;
XX
XX
XX 20-MAY-2004 (first entry)
XX
XX Human ovarian cancer DNA marker #21548.
DE

```

xx Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.

PS Disclosure; SEQ ID NO 21548; 106pp; English.

The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment corresponding to the marker. The presence of protein or protein

fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention.

Sequence 7764 BP: 2342 A: 1534 C: 1892 G: 1979 T: 0 U: 17 Other: 0

Alignment Scores:	
Pred. No.:	0.000171
Score:	164.00
Percent Similarity:	33.87%
Best Local Similarity:	20.02%
Query Match:	5.10%
DR:	5
Length:	7764
Matches:	162
Conservative:	112
Mismatches:	252
Indels:	283
Gaps:	37

SEQ3 D1219 COPY 650 1270 (1-621) x ADL633336 (1-7764)

Qy	2	AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyHisVal	21
		: : : : :	
Db	3342	GCAGCACAGTCGTCTATCTTTGACTCTGACCTGGNACCCCGAGAATGACTTGCAGGACAA	3401
Qy	22	GlulysIleIysIleGluSerCysSerGluAArgThrLysIlePheArgLeuTyrSerVal	41
		: : : : : : : : : :	
Db	3402	GCCCGAGCCGATAGAAATGGTTCAGAGAAGACGAGTAAATATTTACCGCTAGTATTACAAAG	3461
Qy	42	CysThrValGluGluLysAla	49
		: : : : :	
Db	3462	GGGACTGTGGAGAGAGAGATCATAGAACGGGCCAAAGAGAGTGTATTAGATCATCTG	3521
Qy	50	IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn	64
		: : : : : : : : : :	
Db	3522	GTGATTTCAGCGCATGGACACCACTGGCCGAGCATCTCGNAAAACAACACAGGAAGTCC	3581
Qy	65	-----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
		: : : : :	
Db	3582	AACTCAAAATCCTTTTAATAAGAAGAGCTGCACAGCTATTTTGAAATTTGGAGCAGAGGAT	3641
Qy	81	LeuPheAspLysIleuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
		: : : : : : : : : :	
Db	3642	CTCTTCAAGAAGATCGAA	3671
Qy	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
		: : : : :	
Db	3672	GAACCTCAGGAATGGAT-----ATAGATCAAAATTTTGGCTGGCTGAA	3716
Qy	121	GlyGluGluAsnGluVallyLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140
		: : : : :	
Db	3717	ACGAGAGAGAATCAAGTGTCAACAGTGCAACAGATGCAACTTCTA--TCACAGTTTAAG	3773
Qy	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
		: : : : : : : : : :	
Db	3774	GTTGCCAACTTTGCCAACATGGGAAGATGAAGAGAGCTA-----GAAGAGCGT	3821
Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer	180
		: : : : :	
Db	3822	CCTCACAAGACTGGGATGAGATCAIT-----GGGAGGAGATCA	3848
Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu	194
		: : : : :	
Db	3849	-----CCAGAGGAACAAGGAAAAAAGTAGAGGAGGAAGACGCCGAGAGGAGCTAGAA	3902
Qy	195	-----GlySerGluAlaSerProLysThrGly	203
		: : : : : : : : : :	
Db	3903	GAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAGAGCTCAGACAAATGACAGT	3962

3522 GTGATTGAGCGCATGGACACACTGGCCGCGACGATCTCGGAAACAACTCAGGAGGTCC 3581
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db AACTCAAACTCTTTTAAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGAT 3641
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db CTCTTCAAGAACTGGAA-----GGGAGGAAATCA 3671
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db GAACCTCAGAAATGGAT-----ATAGATGAATTTTCCGGTGGCTGAA 3716
QY 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db ACGAGAGAGAAATGAAGTGTCAACAAGTCAACAGTCAACAGATGAATCTCTA---TCACAGATTAAAG 3773
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db GTTGCCCAACTTTGCAACAATGGAAGATGAAGAAGAGCTA-----GAAGAGCGT 3821
QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer 180
Db CCTCACAAGACTGGGATGGATCAAT----- 3848
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu----- 194
Db -----CCAGGACAAAGGAAAAAGTAGAGGAGGAGGAGGAGGAGGAGCTAGAA 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
Db GAAATTTATATGCTGCTCGAATTCGGAGTTCCATAAAGGCTCAGACAAATGACAGT 3962
QY 204 AspGlyClyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
Db GACTCTGACACTGAGTCTAAGAGCGCAGCCGACGAGATCTCTGCTTCTGAGAGTGAAACG 4022
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db -----TCTGATGATGACAGAAG----- 4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db -----CCAAAGCGCAGAGCGCTCCGAGGAGTGTGGGAGGAGGACCTCTGTCGAGGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheClyLeuTyrSerMetGlySer 278
Db TTACTGATCGACAGATCCGAAGGTTTCATCAAGGCTTATAAGAGTTT----- 4145
QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db -----GGTCTCCTCTTGAACGGCTGGAGTGTAGCAGTGAT---GCTGAGCTG 4193
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 4229
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db GAACCTGATCCAC-----AACAGCTGTGTCTCAGCAATGAGGAATATGAAGAGCAGCT 4283
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db AAGAAAATCCAGCGGGAAGGAAAGGNCAGGGAAGG---AGAGGTCCAAACATCAAG 4340
QY 357 LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db ATATCCGGAGTT---CAGGTTAATGTGAATCCATTATCCAAATGAAGAGGAGTT--- 4394
QY 377 ArgAlaAsnSerGlnLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db -----GAGATGCTGCATAAATCTATCCCTGTGAGCCCTGAAGAAAAA 4442

QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db TACTGCTTAACCTGCTGCTGTCACAAAGCTGCACATTTTGTATGTAGAGTGGGGTGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer--- 422
Db GATTCTCGCGCTGTTGCTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATTAA 4562
QY 422 sGlnSerValSerThrLysLeuValAsn----- 432
Db -----ACCAGAGCTTAATTAATCTGACAAAATTTCTCCGCTGGAGACA 4610
QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db GATAAAAGAGCTCAGGGAGAGCAGCTACAGACCGGAGCGATTTACTTGTGTAAGCTGCTC 4670
QY 451 ArgAsnSerSerGlu-----AspGluCysMetThrGluLys 463
Db AGAAAGGCTCTGGAGAAAGAGGGGCTGTGACAGGTGGGAGGAGGCCAAATTAAGAGAG 4730
QY 464 Arg----- 464
Db CGGAAGCTCGGTAAGAGGAACAAAGTCCCGAGGTGAAAGAGGAGCATGGAATT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGAGGAGTAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db GCGTTGCGAAAAAGTCCCAATGAAAAAACAAGAG---AAGAAGAGAGAACAGGAGAAC 4907
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db AAGAGAAACAAATAGTTCTAGGAAGAACAAAGAGGAGGAGCAAGAAAGAAAGACTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db AAGATAAGAAAGAGAGAGCTTAAAGTGGTGATGTCCTCAATCTTCGAGTAATCAAGCGA 5027
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db TCTCAGGCTCTGCTCCATATTACAGCAGGAGGAGAACCTGTCTCCCATTTGGAGAGGATGAG 5087
QY 539 ----- 539
Db GATGATGATCTGGACCGAGAGACATTCAGCATATGTAAGAGAGGATGAGCCCGTGAAA 5147
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db AAGGCACCTGAACAGCTCGACAAACCTGACAAAGGGGCTCAACGTGCAAGAAACAGCTGAA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db CACACCCGAACTGCTGCTGCTGAAATTCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
QY 564 ----- 564
Db TCAGATCAGAGCAGCATCAAACTCTCGAGGAGGAACCTATGAGATTTTGTTCCAAGTTT 5327
QY 565 -----LeuGluGlnMetHisAspValAlaLysLysLysLeuAla 577
Db ACAGAAATTTGATGCTCGAAAACTGCATAGTTATACAGATGCTCATAGAAGAGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
Db CAAGAAGAGAGAG-----GAGCAAAAGAGAAAGACGAC 5420
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db -----ACTGGGGGTAAAGAACCATTTTCTGCCAGGCGCTCAGGCTCCAGCGG 5471

QY 614 -----IleSerGlnAen 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492

RESULT 12

ADK60759

ID ADK60759 standard; DNA; 7764 BP.

AC AC

XX ADK60759;

DT 06-MAY-2004 (first entry)

XX XX

DE Angiogenesis differentially expressed gene GS-N32.

XX ds; vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
KW antinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
KW angiogenesis; endothelial cell diagnosis; tumor vascularization;
KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;
KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
KW ischemia; angina; myocardial infarction; chronic heart disease;
KW cardiac congestion; macular degeneration; osteoporosis.

XX Homo sapiens.

OS

XX FR2836686-A1.

PN

XX 05-SEP-2003.

XX

PF 04-MAR-2002; 2002FR-00002717.

XX

PR 04-MAR-2002; 2002FR-00002717.

XX

PA (GENE-) GENE SIGNAL.

PA (ALMA/) AL MAHMOOD S.

XX

PI Colin S, Schneider C, Al Mahmood S;

XX WPI; 2004-013911/02.

DR P-PSDB; ADK60509.

XX

PT Compositions containing nucleic acid or polypeptide differentially
PT expressed in angiogenesis are useful to diagnose, prognosis and treat
PT angiogenic disorders including tumor vascularization and heart disease.

XX Claim 2; SEQ ID NO 33; 405pp; French.

XX

CC The invention relates to a novel pharmaceutical composition active on
CC angiogenesis comprising an endothelial cell nucleic acid whose expression
CC is induced by an angiogenic factor and inhibited by an angiotatic agent
CC or its complement or fragment, a polypeptide sequence encoded by the
CC nucleic acid or its fragment, a molecule capable of inhibiting expression
CC of the nucleic acid or a molecule which binds to the polypeptide
CC sequence. The invention is used to diagnose, prognosis or treat an
CC angiogenic disorder in a mammal, particularly a human. The disorder is
CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
CC endometriosis associated with neovascularization, restenosis due to
CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
CC myocardial infarction, chronic heart disease, cardiac congestion or
CC macular degeneration due to age or osteoporosis. This sequence
CC corresponds to a differentially expressed DNA used in the composition of
CC the invention.

XX Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

SQ

Alignment Scores:

Pred. No.: 0.000171 Length: 7764

Score: 164.00 Matches: 162

Percent Similarity: 33.87% Conservative: 112

Best Local Similarity: 20.02%		Mismatches: 252
Query Match: 5.10%	Indels: 283	
DB: 12	Gaps: 37	
SEQ3_D1219_COPY_650_1270 (1-621) x ADK60759 (1-7764)		
QY	2	AlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisAspVallyshHisVal 21
Db	3342	GGGACACAGTCGTCATCTTTGACTCTGACTGGACCCCGCAGATGACTTGCAGACCAA 3401
QY	22	GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db	3402	GGCCGAGCGCATAGAAATTCGTCAGAAGACGAGGTAATATTACCGCTAGTTACAAAG 3461
QY	42	CysThrValGluGluLysAla 49
Db	3462	GGGACTGTGGAGGAGGAGATCATAGACGGGCCAAAAGAGATGTTATTAGATCATCTG 3521
QY	50	IleLeuAlaArgGlnAenMetArgGlnAenLysAlaValGluAen 64
Db	3522	GTGATTTCAGCGCATGGACACCACTGCGCGGACGATCTCTGAAAACAACTCAGGAAGTCC 3581
QY	65	LeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db	3582	AACTCAAAATCCTTTTAATAAGAACGAGCTGACAGCTATTTTGAATTTTGAGCAGAGGAT 3641
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db	3642	CTCTTCAAAAGAACTGGA 3671
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db	3672	GAACCTCAGGAATGGAT 3716
QY	121	GlyGluGluAenGluValLysLeuCysLeuLeuLeuAlaLysHisAlaGlnGlyThr 140
Db	3717	ACGAGAGAGAATGAAGTGTCAACAAGTGCACAGATGAATCTCTA--TCACAGTTTAAAG 3773
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db	3774	GTTGCCAACTTGCACAAATGGAAGATGAAGAAGAGCTA-----GAAGAGCGT 3821
QY	161	ProAenIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db	3822	CCTCACAAGGACTGGGATGAGATCAAT 3848
QY	181	AspThrProGlnArgAenArgLysArgValGlnTyrPheGlu 194
Db	3849	CCAGAGAACAAAGGAAAAGTAGAGGAGGAGAGCGGAGAGGAGGAGTAGAA 3902
QY	195	GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAGAGCTCAGACAAATGACAGT 203
Db	3903	GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAGAGCTCAGACAAATGACAGT 3962
QY	204	AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
Db	3963	GACTCTGACACTGAGTCTTAAGAGGAGCGGCCAGAGATCCTCTGCTCTGAGAGTGAACG 4022
QY	224	ThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db	4023	GAAGAC-----TCTGATGATGACAGAAG-----4046
QY	244	LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db	4047	-----CCAAAGCGCAGAGGCGCTCCGAGAGGTGTGCGGAAGGAGCACCTCTGTGGAGGGA 4097
QY	264	ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db	4098	TTTACTGTATGCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT-----4145
QY	279	HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db	4146	-----GGTCTCCCTCTTGAACGGCTGGAGTGTCTAGCAGCTGAT--GCTGAGCTG 4193

QY 297 ProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetalalysLeuCys 316
 Db 4194 GTAGATAAGTCGGTGGCAGATCTG
 QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPhelLeuGluTyrVal 336
 Db 4230 GAATCTGATCCAC
 QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
 Db 4284 AAGAATAATCCAGCGAGGAAAGGACAGGGAAGG--AGAGGTCCACATCAAG 4340
 QY 357 LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 4341 ATATCGGAGTT--CAGGTTAATGTCAATCATTCATCAACATCAAGAGGAGTT-- 4394
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 4395 -----GAGATGCTGCATAAATCTATCCCTGTGGACCTGAAGAAAAA 4442
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 Db 4443 TATGCTTAACCTGCTGTCGAAGCTGCACATTTTGTAGTGGGGGTGGAGAT 4502
 QY 415 AspCys-----PheGly-ThrAsnSer-----Ly 422
 Db 4503 GATTCCTCGCTGTTGGGATTTATGAACATGGCTATGGAACCTGGAGTTAATTAA 4562
 QY 422 sGlnSerValValSerThrLysLeuValAsn----- 432
 Db 4563 ACAG-----ACCAGAGCTTAAATTAACAGCAAAATTCGCGGTGGAGACA 4610
 QY 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
 Db 4611 GATAAAAGCTCAGGGAGAGCTACAGACCGGAGCGATTACTTGTGAAGCTGCTC 4670
 QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
 Db 4671 AGAAGGGTCTGGAGAAGAGGGGGCTGCACAGGTGGGGAGGAGGCCAAATTAAGAG 4730
 QY 464 Arg----- 464
 Db 4731 CGGAAGCTCGGTAAAGAGGAAAAAAGTCCCGAGCTGAAGAGGAGCATGGAATT 4790
 QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
 Db 4791 GAGCTTTCATCTCTAGGCATTCAGATAATCCATCAGAGGGGAGAGTGAAGATGAT 4850
 QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
 Db 4851 GGCTTGAAAAAAGTCCCAATGAAAAAAGGAGAG--AAGAAAGAGAACAGAGAGAC 4907
 QY 501 LysMetGluLeuLeuLeuMetTyrAlaAspLys----- 511
 Db 4908 AAGGAGAAACAAATGATGTTCTAGGAAAGACAAAGAGGGGACAGGAAAGAAAGAGTCA 4967
 QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
 Db 4968 AAGATAAAGAAAGAGAGCCCTAAAGTGGTATGTCGCAATCTTCGAGTAAATCAAGCGGA 5027
 QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
 Db 5028 TCTCAGGGTCTCTCCATATTACAGCAGGAGGAGTGAACCTCTGCCCATTCGAGGATGAG 5087
 QY 539 ----- 539
 Db 5088 GATGATGATCTGGACAGGAGACATTCAGCATATGTAAGCAGAGGATGAGGCCCGGTGAAA 5147
 QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
 Db 5148 AAGGCATGAAACAGCTCGACAAACCTGACAAAGGGGCTCAACGTGCAAGAACAGCTGGAA 5207

QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
 Db 5208 CACACCCGGAACCTGCTGCTGAAATCGGAGACCGGATAGCCGAGTCCCTTAAAGCCTAC 5267
 QY 564 ----- 564
 Db 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGGATTTTGTTCCAAGTTT 5327
 QY 565 -----LeuGluGlnMetHisAspValAlaLysLysLysLeuAla 577
 Db 5328 ACAGAAATTTGATGCTCGAAAACTGCATAAGTTATACAAGATGCTCATAGAAGAGTCT 5387
 QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
 Db 5388 CAAGAAAGAGAG-----GAGCAAAAGAGAAAGACGAC 5420
 QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
 Db 5421 GTG-----ACTGGGGTAAAGAACCCATTTCTCCAGAGCGCTCAGGCTCAGCGG 5471
 QY 614 -----IleSerGlnAsn 617
 Db 5472 GACTCTCTGATATCTCAGTCC 5492
 RESULT 13
 ADP73081
 ID ADP73081 standard; DNA; 7764 BP.
 XX
 AC ADP73081;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Angiogenesis inhibitor human DNA sequence, GS-N32.
 XX
 KW Inhibitor; angiogenesis; antisense nucleic acid; immunisation;
 KW angiogenic disorder; antiangiogenic; angiogenesis stimulator; cytostatic;
 KW dermatological; antiarthritic; antirheumatic; antiinflammatory;
 KW vasotropic; hypotensive; ophthalmological; antipsoriatic; cardiant;
 KW gene therapy; antisense gene therapy; tumour vascularisation;
 KW retinopathies; rheumatoid arthritis; Crohn's disease; atherosclerosis;
 KW ovarian hyperstimulation; psoriasis; endometriosis; restenosis;
 KW tissue granulation; peripheral vascular disorder; hypertension;
 KW vascular inflammation; Raynaud's disease; aneurism; arterial restenosis;
 KW thrombophlebitis; lymphadenopathy; lymphedema; ischaemia; angina;
 KW myocardial infarction; chronic heart disease; congestive heart disease;
 KW macular degeneration; human; ds; gene.
 XX
 OS Homo sapiens.
 XX
 PN FR2843753-A1.
 XX
 PD 27-FEB-2004.
 XX
 PF 20-JUN-2003; 2003FR-00007507.
 XX
 PR 04-MAR-2002; 2002FR-00002717.
 XX
 XX (GENE/) GENE S.
 XX (ALMS/) AL M S.
 XX
 PI Colin S, Schneider C, Al MS;
 XX
 XX WPI; 2004-216677/21.
 DR P-PSDB; ADP73132.
 XX
 XX Antisense nucleic molecule useful as inhibitor of angiogenesis in the
 PT treatment of angiogenic disorders, e.g., rheumatoid arthritis,
 XX atherosclerosis and endometriosis.
 PS Claim 1; SEQ ID NO 33; 404pp; French.
 XX
 CC The invention relates to a novel inhibitor of angiogenesis comprising an
 active substance chosen from at least one of a nucleic acid molecule, an

CC antisense nucleic acid molecule, a polypeptide or an antibody. The
 CC invention further comprises: an antisense nucleic acid sequence chosen
 CC from any of the sequences provided in the specification; preparation of
 CC the antibody comprising in vivo or in vitro immunisation of an
 CC immunocompetent animal cell, preferably of a vertebrate and most
 CC preferably of a mammal, with at least one of the polypeptide sequences
 CC chosen from a sequence provided in the specification; a mammalian
 CC expression vector comprising at least one antisense sequence chosen from
 CC an antisense nucleic acid provided in the specification; preparation of a
 CC genetically modified cell, that over- or under-expresses a gene
 CC implicated in an angiogenic disorder, comprising inserting the vector
 CC from above into a mammalian cell; a genetically modified cell that over-
 CC expresses or under-expresses at least one gene involved in angiogenesis
 CC by a nucleic acid sequence chosen from any of ADP73049 to ADP73338, as in
 CC the antisense nucleic acid selection of above, or a fragment of any of
 CC these; preparation of a cell line that stably expresses an expression
 CC vector. The angiogenesis inhibitor has the following activities:
 CC antiangiogenic, angiogenesis stimulator, angiogenesis inhibitor,
 CC cytotatic, dermatological, antiarthritic, antirheumatic,
 CC antinflamatory, vasotropic, hypotensive, ophthalmological,
 CC antipsoriatic, and cardiant. The novel nucleic acid molecules of the
 CC invention may be used to treat disorders in gene therapy and antisense
 CC of the therapeutic compositions are useful in treating a disorder of
 CC angiogenesis chosen from: tumour vascularisation, retinopathies,
 CC rheumatoid arthritis, Crohn's disease, atherosclerosis, ovarian
 CC hyperstimulation, psoriasis, endometriosis associated with
 CC neovascularisation, restenosis due to balloon angioplasty, tissue
 CC granulation due to scarification, peripheral vascular disorders,
 CC hypertension, vascular inflammation, Raynaud's disease, aneurism,
 CC arterial restenosis, thrombophlebitis, lymphadenopathy, lymphedema,
 CC ischaemia, angina, myocardial infarction, chronic heart disease,
 CC congestive heart disease, macular degeneration linked to age and
 CC osteoporosis. This polynucleotide represents a human DNA sequence for the
 CC creation of an angiogenesis inhibitor of the invention.

XX Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000171 Length: 7764
 Score: 164.00 Matches: 162
 Percent Similarity: 33.87% Conservative: 112
 Best Local Similarity: 20.02% Mismatches: 252
 Query Match: 5.10% Indels: 283
 DB: 12 Gaps: 37

SBQ3_D1219_COPY_650_1270 (1-621) x ADP73081 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
 DB 3342 GCGGACACAGTGTCTATCTTGAATCTGACTGGAACCCCGCAAGTACTGACGGCACAA 3401
 QY 22 GluLysIleLeuIleGluSerCysSerGluArgThrLysIlePheArgLeuThrVal 41
 DB 3402 GCCCGAGCCGATAGAAATTCGTGAGAGAGAGAGTAAATATTACCGCTTAGTTACAAG 3461
 QY 42 CysThrValGluGluLysAla-----Leu 49
 DB 3462 GGGACTGTGGAGGAGGAGATCATAGACGGGCCAAAGAAAGATGTTATAGATCATCTG 3521
 QY 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
 DB 3522 GTGATTACGGCGATGGACACCACTGCGCGAGCATCTCGAATAAACCACTCAGGAAGGTCC 3581
 QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 DB 3582 AACTCAAAATCTTTTAAAGAAAGAGCTGACACGCTATTTTGAATTTGGAGCAGGAT 3641
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 3642 CTCTTCAAGAACTGGAA-----GGGAGGAATCA 3671
 QY 101 GluThrSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120

DB 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTCGGTGGCTGAA 3716
 QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 DB 3717 ACGAGAGAGATGAAGTGTCAACAGTCCACAGATGAATCTCTA---TCACAGTTTAA 3773
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer 160
 DB 3774 GTTGCCAACTTTGCCAACAAATGCAAGATGAAGAAGACTA-----GAAGAGCGT 3821
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer 180
 DB 3822 CCTCACAAGGACTGGGATGAGATCATTT-----3848
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 DB 3849 -----CCAGAGGAACAAGGAAGAAAGTAGAGGAGGAAGAGCGCAGAGGAGTAGAA 3902
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 DB 3903 GAAATTTATATGTCCTCGAATTCGGAGTTCCTCAATAAAGGCTCAGACAAATCAGACT 3962
 QY 204 AspGlyGlyAsnAlaLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
 DB 3963 GACTCTGACACTGAGTCTTAAGAGGCGAGCCAGAGATCCTCTGCTTCGAGAGTGAACG 4022
 QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 DB 4023 GNAGAC-----TCTGATGATGACAAAG-----4046
 QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 DB 4047 -----CCAAAGCGCAGAGGCGTCCGAGGAGTGTGCGGAGGACCTCTCTGGAGGA 4097
 QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 DB 4098 TTTACTGATGACAGATCCGAAGTTCATCAAGCTTATAAGAGTTT-----4145
 QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 DB 4146 -----GGTCTCCCTTGAACGGCTGGAGTGTAGCAGCTGAT---GCTGAGCTG 4193
 QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 DB 4194 GTAGATAAGTCGGTGCAGATCTG-----AAGCGCCTGGGT 4229
 QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 DB 4230 GAACGTGATCCAC-----AACAGCTGTGTGCAGCAATGCAGGAATATGAAGAGCAGCTG 4283
 QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
 DB 4284 AAAGAAATGCCAGCGAGGAAAGACCCAGGAAAGG---AGAGGTCCAACAATCAAG 4340
 QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 DB 4341 ATATCCGGAGTT---CAGGTTAATGTGAATCCATTATTCACATCAAGAGAGGAGTT---4394
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 DB 4395 -----GAGATGCTGCAATAATCTATCCCTGTGGAGCCCTGAGAGAAAAA 4442
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 DB 4443 TACTCTTAACCTGTCGTCAAAGCTGCACATTTTGTAGTAGAGTGGGGGTGGAAGAT 4502
 QY 415 AspCys-----PheGly-ThrAsnSer-----ly 422
 DB 4503 GATTCCTCGCCTGTTGCTGGGAGTTTATGAACATGGCTATGGAAGACTGGGAGTTAATAA 4562
 QY 422 sGlnSerValSerThrLysLeuValAsn-----432


```

QY 203 yAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVa 223
Db 1478 CCATGGCCCTGGACTTCGGTTTTCGCAGCCGGATCTCGGAGCACCTGCTGATGTTGATGT 1537
QY 223 lThrAspPro-----ProValAspAa 230
Db 1538 GCTTTTCCCACTGCTGGTGGAGCTTGTGGCAGGCCCGCCAGCAACCACTGGGAATAGAAGA 1597
QY 230 pAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----246
Db 1598 CAAGGATGACAGCCAGTCCAGCCCAAGATGAGCTGCGAGCAAGCAGTCCAAAGGCCCTGGA 1657
QY 247 -----ProLysValIleThrLeuGlnSerSerCy 256
Db 1658 GGAGAGGTACCATAGTTATCTCTCCACTTCCACAGCAGGAGCGGGCCAGAGTCCCCC 1717
QY 256 sLysSer-----SerGlyThrAspGlyThrLeuAsp-----266
Db 1718 TCGCAGCCTGGCCACTGAAGAAGAGAGCCCGCCAGGGCCCGAGGGCAGCCCGAGTGGAA 1777
QY 267 -----GlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSe 281
Db 1778 GGAGGCAGAGGAGCTTGGGGAGGACTCTGCAGCCAGCCCTCAGCCTGCAGCTGTCCCTCCA 1837
QY 281 r-GlyIleProGluAspMetLeuAlaSerGln-----291
Db 1838 GAGGCGATCCACAGAGCCTGTGGCTCCCGCCAGAGCAGCTCTCAGAGGCTGCATAAAGGC 1897
QY 292 -----AspTrpGlyLysIleProAspGluSerGlnArgArgLeuHisThrValL 308
Db 1898 CATGGAAGAGGAGTGGCCCAAGTACTC-GAGCAAGACAGAGGACCACTGCTGGAATCCA 1956
QY 308 euLysProLysMet-----AlaLysLeuCysGln-----ValL 319
Db 1957 AGCAAGAGAGATGCAGCAACTCTCGGAGAGAGCTGTGCCAAGAGAGGAGGAGGAGGAGATCC 2016
QY 319 euHisLeu-----SerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrV 336
Db 2017 TCGGGCTTCCACAGCAGAAAGAGCAATCTCTCAGTTCTTGGAGGGAGCGCTGCAGAAAG 2076
QY 336 alIleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleA 356
Db 2077 CCATTGAG-----GAGGAGGAGCCCGGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2121
QY 356 laLeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuV 376
Db 2122 GGTATCTCTGGCTCCGAGCT---CAGGTCAGTCCAGCAGCAGCAGCAGCAGCAGGAGGAGGAG 2178
QY 376 alArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrS 396
Db 2179 TCAGGGCTGAGCAAGAGGCTTCCCTGCAGAAACTGAGAGAGAGAGTGGAG-----2228
QY 396 erIleLeuSerCysMetLys-----SerLeuPheLeuGluHisThrGlnGlyL 412
Db 2229 -----TCTCAACAGAGGCTGAGAGGGCCAGCTTGGAAACAGAAAAATAGGCAATGC 2280
QY 412 eu-----GlnPheAspCysPheGlyThrAsnSerLysGlnSerValValS 427
Db 2281 TGAGCAGCTCAAGGAAGAGATGAGGCT---TCGGAGAGAGCAGCAGCAGCTGCCCTGA 2337
QY 427 erThr-----LysLeuValAsnGluSerLeuSerGlyAlaThrValA 441
Db 2338 ATGCTCAAAAGGAGAGGCTCTCAGCAGCTGAGGGAGCAGCTGGAAGGGAG-----A 2391
QY 441 rgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMetT 461
Db 2392 GGAAGAAGAGCTGTGGCAAGCCTCGAGAAGGAGAGCAGTGTCTGAG-----C 2436
QY 461 hrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIle-----GluL 477
Db 2437 TGGAGCGGCTCTCTCTCTATTGGAGCCCAAGCAGCGGAGGAGTGGTCTCCAGCCTCCAGA 2496

```

```

QY 477 yThrIleSerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluH 497
Db 2497 AGAAGATACAGGAAGCTCAACAGAAAGAGAGGCCCGCAGCTGCAGAAGTCCCTTGGGCAAG 2556
QY 497 isGluGluLys-----LysMetGluLeuLeuA 506
Db 2557 TGGAGCAGCAGAGTTCCACCAAGTCTTATCAGCTGGCTGGGTATGAGCAGCAGAGCTCAGCA 2616
QY 506 snMetTyrAlaAspLysLysGlnLysLeuGlu-----516
Db 2617 GTCTCTCGCAGAGAGCGCCAGGAAGTGAAGGGGAGCATGAGAGGAGGTTGGACAAGA 2676
QY 517 -----ThrSerLysSerValGluAlaAlaValI 526
Db 2677 TGAAGGAGGAGCACCAGCAAGTGATGGTGAAGCCAGAGCAGATGAAGCTGAGGAGA 2736
QY 526 leArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuAspHisA 546
Db 2737 GGAAGCAGCGGCTGAGCTTCTGGGGCAGCTGACCGGAGAGCTGGAGCGCCCTTGACAGGG 2796
QY 546 snTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuG 566
Db 2797 CCCATGAACGAGAACTGGAGACTGTGAGCAGCAGCAACACAAAGCGCTCTTGAGGACTTGC 2856
QY 566 luGlnMetHisAspValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnA 586
Db 2857 GCGCGCGGCACAGGAGCAGGAAGGAGCTCCAGGATTTAGAGTTGGACCTTGAACCA 2916
QY 586 rgIleLysSerTrpAlaAlaLysLeu 594
Db 2917 GAGTAAAGATGTCAAGGCCAGATTG 2942

```

RESULT 15
 ADS48467
 ID ADS48467 standard; cDNA; 3045 BP.
 XX
 AC ADS48467;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polynucleotide #3210.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX
 OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 26897; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 3045 BP; 1039 A; 712 C; 819 G; 475 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.000105 Length: 3045
 Score: 160.00 Matches: 131
 Percent Similarity: 37.00% Conservative: 111
 Best Local Similarity: 20.03% Mismatches: 237
 Query Match: 4.97% Indels: 176
 DB: 13 Gaps: 29

SEQ3_D1219_COPY_650_1270 (1-621) x ADS48467 (1-3045)

QY 17 AspValHisValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePhe 36
 DB 673 GATCGAGCAITGGCGGCAAGTGCACACCGATCTCAAAAAGAGAGGAGATTCTCTATCT 732
 QY 37 ArgLeuTyrSerValCysThrValGluGluLysAlaLeuLysIleLysIleGluSerMet 56
 DB 733 CGTTG-----CAGAGAGAGCT-----ACTGCCAAACCCGAG 765
 QY 57 ArgGlnAsnLysAlaValGluLysAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrp 76
 DB 766 CAGGCCGACCATGTCGGAATCTTAAACCGAACGCTGAAC----- 807
 QY 77 GlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSer 96
 DB 808 -----GATATCATGTGATACCGCGCAGCGCTTGGCTTCTATCAAA 849
 QY 97 GlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeu 116
 DB 850 GGGCGGCCACCTCATGACATAGGCGCAACATCTTCAAGTCTTCGATACATATTG 909
 QY 117 SerSerLysGlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLys--- 135
 DB 910 AACCTGTAC-----GAAGACACCGACCAACGACTTGTTCAAACCGAAGAAGCCAGGAC 963
 QY 136 HisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGluAspHisIle---Lys 154
 DB 964 AAGGCAGAACAGCGGCCCAAGGAATCCGCCACACGATCCCGCAGTGAACAATCCAGAG 1023
 QY 155 LeuSerAspGluGluSer-----ProAsn 162
 DB 1024 CTCAGTCTGAAACACACAGCTTAAAGACAGAGAGATGACGCAAAAAGAGCCGAC 1083
 QY 163 IlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182

DB 1084 GCCTTTTACCAAACTC-----AAGAAAAAGGAGTGGCTTGGCCGACGAGGAT 1134
 QY 183 Pro-----GlnArgAsnArgLysArgValGlnTyrPheGluGly 195
 DB 1135 CCATCGCGCTCACTGCCCTCGACCAAGAGTCTTGAAGACCAACACAGATCGCTGGAG 1194
 QY 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArg-----LysLysAla 214
 DB 1195 TCTGAAAGAATGCCCTGAACCAAGAAATCACTAATCTCAACAAATCGATCGAAGAGCC 1254
 QY 215 SerAspAspValThrAspProArgVal-----ThrAspProValAspAsp 230
 DB 1255 AATGTGGCTCAAGGACATGAAGCTTATGACAGTGGAGAGGATTTCTTGGTCGATCTA 1314
 QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----ProLys 248
 DB 1315 CAACAAGCTCAGAA-----GGGATATCAAAAGCTTGNACCAACAG 1356
 QY 249 ValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsn 268
 DB 1357 GTTCTGACCTCAAGCAAAATTTGGCGAGCAAGGCCAGCGCGCGCGCA-----TCT 1409
 QY 269 AspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyLeuProGluAspMetLeu 288
 DB 1410 GAAGGATATTCACTCTTTCGATTGGAGAACAAAGTCTTTGGAGACCAAGCAACCTGT 1469
 QY 289 AlaserGlnAspTrp-----GlyLys 295
 DB 1470 GCAGCGGAAAGAGGAGTCTTGCAGCAGCAGCTTTCTCAGACGACGAGGCTGGAAAA 1529
 QY 296 IleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu 315
 DB 1530 GGTGGAGACGACCTCAAGAA-----CACGGCTCGCAAAAGATGATTT----- 1574
 QY 316 CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyr 335
 DB 1575 -----GGAGACTCAAAAGAAATGA 1592
 QY 336 ValIleGluAsn-HisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIle 355
 DB 1593 GTGCTCAAGGCAAGAAAGTCTCTGGAGGAGAG----- 1626
 QY 355 eAlaLeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLe 375
 DB 1627 -----ATCAGCAGGCTCAAGAGAGGAGTTGACTTGGCCAGGAAGAAATTCGC 1673
 QY 375 u-----ValArgAlaAsnSerGluLeuAlaPheLysCysSerArgva 389
 DB 1674 TGCGCCAACAACTCGTCGAGATCAAAACCGGAACCCCACTTGGCCGAAAGCATCACTCGCT 1733
 QY 389 lGluVal-----AspTyrIleTyr 395
 DB 1734 CGAACTCATCATGAGGACCAAAACCTCACTACCGTCGACCGCAGCTTCGTCCA 1793
 QY 395 rSerIleLeuSerCysMetLysSerLeu-----PheLeuGluHisThrGlnIle 411
 DB 1794 AGGACTGTTGGCAAGTTGAAGGACGCGCAGGAGACACAAAGACTTGGATCCAGTAGGA 1853
 QY 411 yLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuVa 431
 DB 1854 TGTCTTCGTGAATGATCGCTCGC-----CTTCT 1880
 QY 431 lAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn----- 446
 DB 1881 CAATCGTCTTGAAG---GAGCTGTCTTTGATGAGAGGACAGGAGATGTTCTGAC 1937
 QY 447 -----ThrLysSerMetArgAsnSerSerGluAspGluGlu-----Cy 459
 DB 1938 TTTTGCAAAATGAGTTGAATGAGGCTCTGAGAGAGGCGAAGGAGGACCGCGAAACAAAGCGC 1997
 QY 459 sMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGlyThrIle 479


```

Db 2760 GCGGACACAGTCGTCACTCTTGAAGTCTGAGTCCGAAACCCAGAGTACTTGCAGGCACAA 2819
Qy 22 GluLysIleLeuValSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2820 GCCCGAGCCATAGATTGGTCAGAGAGAGCAGGTAAATATTACCGCTTAGTACAAAG 2879
Qy 42 CysThrValGluGluLysAla-----Leu 49
Db 2880 GGGAGCTGGAGGAGGAGATCATAGAACGGGCGCAAAAGAGAGATGGTATTAGATCATCTG 2939
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
Db 2940 GTGATTACAGCGATGGACACCACTGGCGGAGCATCTCGGAACCAACTCAGAGAGTCC 2999
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTdpGlyAlaSerTyr 80
Db 3000 AACTCAATCCTTTTAAATAAGAGAGCTGACAGCTATTATTGAAATTTGGAGCAGAGAT 3059
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3060 CTCTTCAAGAACTCGAA-----GGGAGGAATCA 3089
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3090 GAACCTCAGGAATGGAT-----ATAGATGAATTTTGGCGTTGGCTGAA 3134
Qy 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3135 ACGAGAGAGATGAAGTGTCAACAGTGCACAGATGAATCTTA---TCACAGITTAAG 3191
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3192 GTTGCCAACTTTGCCAACATGAAGTGAAGTGAAGAGAGCTA-----GAAGAGCGT 3239
Qy 161 ProAsnIlePheTyrSerLysLeuLeuGlyGlyLysAsnProMetTyrLysTyrProSer 180
Db 3240 CCTCACAAGACTGGATGAGATCAIT-----3266
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
Db 3267 -----CCAGAGAACAAAGAAAGTGAAGAGAGAGAGCGGAGAGAGGTAGAA 3320
Qy 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3321 GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTTAAAGGCTCAGACAAATGACAGT 3380
Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db 3381 GACTCTGACACTGAGTCTTAAGAGGCGAGGCCAGAGATCCTCTGCTTCTGAGAGTGAACG 3440
Qy 224 ThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 3441 GAAGAC-----TCTGATCATGACAAGAG-----3464
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 3465 -----CCAAAGCGCAGAGGGGCTCGAGAGGTGTGCGGAAGGACCTCGTGGAGGGA 3515
Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 3516 TTACTGATGCGAGATCCGAAGGTTTCATCAAGGCTTATAAGAGTT-----3563
Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTyrGlyLysIle 296
Db 3564 -----GGTCTCCTCTTGAACGGCTGGAGTGCATAGCAGCTGAT---GCTGAGCTG 3611
Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 3612 GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCCTGGGT 3647
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 3648 GAACGTGATCCAC-----AACAGCTGTGTGTGCGCATGCGAATACGAGATGAGCAGCTG 3701

```

```

Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 3702 AAAGAAATATCCAGCGAGGAAAGGACGAGGAAAGG---AGAGTCCCAACATCAAG 3758
Qy 357 LeuSerTyrIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 3759 ATATCCGAGTT---CAGTTAATGTAAATCCATTATCCACATGAGAGAGGAGTTT---3812
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db 3813 -----GAGATGCTGCATAAATCTATCCCTGTGGACCCCTGAGAAAAA-----3860
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 3861 TACTCTTAACCTGTCGTGTCAAAGCTGCACATTTTGTATGTAGAGTGGGGGTGGAAGAT 3920
Qy 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
Db 3921 GATTCCTCCCTGTTGCTGGGATTTATGAACATGCTATGGAACACTGGGAGTTAATTAA 3980
Qy 422 sGlnSerValSerThrLysLeuValAsn-----432
Db 3981 ACAG-----ACCGAGCTTAAATTAACCTGACAAAAATCTGCCGCTGGAGACA 4028
Qy 433 ---GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db 4029 GATAAAAGCCTCAGGGGAGCAGCTACAGACCCGAGCGGATTACTTGTGAAGCTGCTC 4088
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4089 AGAAAGGCTTCGGAAGAACAGGGGCTGTGACAGTGGGGAAGAGCCCAATTAAGAAG 4148
Qy 464 Arg-----464
Db 4149 CGGAAGCCTCGGTTAAAGAAACAAAGTGCAGGCTCAAGAGAGGAGCATGGAATT 4208
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrLysSer 480
Db 4209 GAGCTTTTCATCTCTAGGCAATTCAGATAATCCATCAGAAGAGAGGAGAAATGAAATGAT 4268
Qy 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4269 GCTTTGGAAAAAGTCCCAATGAAAAAACAAGAG---AAGAAAGAGAACCAAGGAGAAC 4325
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----511
Db 4326 AAGGAGAAACAAATGAGTTCTAGGAAACACAAAGAGGGGACACAGAAAGAAAGATCA 4385
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----523
Db 4386 AAAGATAAGAAAGAGAGAGCCTAAAAAGTGTGTATGTCCTCAATCTTCAGTAAATCAAAGCGA 4445
Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----539
Db 4446 TCTCAGGCTCTGTCTCATATTACAGCAAGAGTGAACCTGTCCCCATTTGGAGAGGATGAG 4505
Qy 539 -----539
Db 4506 GATGATGATCTGGACAGGAGACATTCAGCATATGTNAGGAGAGGTGAGGCCCGTGAAA 4565
Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 4566 AAGGCACCTGAAACAGCTGCAACAACTGCAAGGGGCTCAACGTGCAAAACAGCTGGAA 4625
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer---564
Db 4626 CACACCCGAACTGCTGCTGTAATAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCTTAC 4685
Qy 564 -----564
Db 4686 TCAGATCAGGAGCACATCAAACTCTGAGGAGGAGAACCTATGATTTTGTTCCAAGTTT 4745

```

```

Qy 565 -----LeuGlnMetHisAspValAlaLysLysLeuAla 577
Db 4746 ACAGAAATTGTAGTCGAAAACTGCATAAGTTATACAAGATGGCTCATAGAAGAGGTCT 4805
Qy 578 GluAspGluAlaCysTrrPileAsnArgIleYsSerTrrPAlaAlaLysLeuLysValCys 597
Db 4806 CAAGAGAGAG-----GAGCAAAAGAAGAAAGACAC 4838
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 4839 GTG-----ACTGGGGGTAAAGAAACCACTTTCGTCCAGAGGCCCTCAGGCTCCAGCGG 4889
Qy 614 -----IleSerGlnAsn 617
Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 2
US-09-949-016-2989
; Sequence 2989, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2989
; LENGTH: 7699
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2989

Alignment Scores:
Pred. No.: 1.37e-07 Length: 7699
Score: 164.00 Matches: 162
Percent Similarity: 33.87% Conservative: 112
Best Local Similarity: 20.02% Mismatches: 252
Query Match: 5.10% Indels: 283
DB: 4 Gaps: 37

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-949-016-2989 (1-7699)
Qy 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
Db 3343 CGCGACACAGTCGTCATCTTTGACTCTGCATGGAAACCCCAAGATGACTTCGAGGCCAA 3402
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3403 GCCCGAGCGCATAGAAATTTGGTCAGAAAGACGAGGTAAATATTACCGCTTAGTTACAAG 3462
Qy 42 CysThrValGluGluLysAla-----Leu 49
Db 3463 GGGACTCTGGAGGAGGAGATCATAGAACGGGCCAAAAGAGATGGTATTAGATCATCTG 3522
Qy 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn----- 64
Db 3523 GTGATTCAGCGCATGGACACACATCGGCGGACGATCTCTGGAAAAACAACCTAGGAAGTCC 3582
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrrPAlaLysValSerTyr 80
Db 3583 AACTCAATCCTTTTATAAAGAGAGCTGCACACTATTTTGAATTTGGACGAGGAT 3642
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100

```

```
QY 422 sGlnSerValSerThrLysLeuValAsn----- 432
Db 4564 ACAG-----ACCCAGAGCTTAAATTAACCTGACAAAATTCGCGGTGGAGACA 4611
QY 433 ---GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
Db 4612 GATAAAAGCCTCAGGGGAAGCAGCTACAGACCGGCGGATTAATTTGTTGAAGCTGCTC 4671
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4672 AGAAAGGCTCTGGAGAGAGGGGGCTGTGACAGGTGGGAGAGGCCAAATTAAGAAG 4731
QY 464 Arg----- 464
Db 4732 CGGAGCCTCGGTAAAGAGAAACAAAGTCCAGGCTGAAAGAGGAGCATGGAATT 4791
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4792 GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGGGAGAGTGAAGATGAT 4851
QY 481 GlyIleLysLeuLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4852 GCCTTGGAAAGTCCATGAAAGAAACAGAG-----AAGAAAGAGAACAGGAGAAC 4908
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4909 AAGGAGAAACAATGAGTTCTAGGAAAGACAAAGAGGGGACAGAGAAAGAAAGATCA 4968
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4969 AAGATAAAGAGAGAGAGCTTAAAGTGTGTGATGTCCTCAATTAATCAAAAGCGA 5028
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 5029 TCTCAGGGTCTGTCCATATTACAGCAGGAGTGAACCTGTCTCCCATTTGGAGAGATGAG 5088
QY 539 ----- 539
Db 5089 GATGATGATCTGGACAGGACATTCAGCATATGTAAGGAGAGGATGAGCCGCTGAAA 5148
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5149 AAGGCACCTGAAACAGCTGCACAAACCTGACAAAGGGGCTCAACCTGCAAGACAGCTGGA 5208
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5209 CACACCCGAACTGCTGCTGAAATCCGAGACCGGATAGCCGAGTGCCTTAAGCCTAC 5268
QY 564 ----- 564
Db 5269 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGGATTTTGTTCCTCAAGTTT 5328
QY 565 -----LeuGluGlnMetHisAspValAlaLysLysLysLeuAla 577
Db 5329 ACAGAATTTTCATCTCGAAAACCTGCAATAAGTTATACAAGATGGCTCATAGAAGAGTCT 5388
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 5389 CAAGAAGAAGAG-----GAGCAAAAGAAAGAAAGACGAC 5421
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5422 GTG-----ACTGGGGTGAAGAAACCATTTTCGTCAGAGGCGCTCAGGCTCCAGCCGG 5472
QY 614 -----IleSerGlnAsn 617
Db 5473 GACTCTCTGATATCTCAGTCC 5493

RESULT 3
US-09-949-016-5001
; Sequence 5001, Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5001
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-5001

Alignment Scores:
Pred. No.: 5,79e-07 Length: 5883
Score: 157.00 Matches: 127
Percent Similarity: 36.62% Conservative: 103
Best Local Similarity: 20.22% Mismatches: 218
Query Match: 4.88% Indels: 181
DB: 4 Gaps: 26

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-949-016-5001 (1-5883)
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2878 CAGAAGCTGCAGCTGGAGAAGGTGACCACCGAGGGCGAAGCTGAAAAGCTG----- 2928
QY 42 CysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 2929 -----GAGGAGGAGCAGATCATCTCTGGAGGACCAAGACTCGCAAGCTGCCAAG--- 2976
QY 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db 2976 ----- 2976
QY 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
Db 2977 -----GAA 2979
QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
Db 2980 AAGAAACTGCTGGAGACAGAAATAGCTGAGTTTCAACCAACCTCACAGAA----- 3030
QY 122 GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
Db 3031 GAGGAGGAGAAATCTAAGAGCCTCCCAAGCTCAAGAACACAGCAT----- 3075
QY 142 SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGlu-SerPr 161
Db 3076 -----GAGGCAATGATCAGTCTGGAAGAGCGCTCCGAGGAGGAGGAGAGCAG 3126
QY 161 AsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAs 181
Db 3127 CCACAGGAGCTGGAGAGAGCCCGCGGAGC-----TGAGG-----GA 3165
QY 181 pThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProly 201
Db 3166 GACTCCACAGACCTCAGCGACCAGA-----TCGCCGAGCTCCAGGCCCGAG 3210
QY 201 sThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspPr 221
Db 3211 ATCGCGGAGCTCAAGATGCACTGCGCCAGAGAGG----- 3247
QY 221 oArgValThrAspProValAspAspGluArgLys----- 234
```


Db 3248 -AGGAGTCCAGCGCCCTCGCCAGAGTGGAGAGAGAGCTGCCAGAGAACATGGCC 3306
QY 235 -----AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal-----11 250
Db 3307 CTCAAGAAGATCCGGGAGCTGGAATCTCAGATCTCTGAATCTCCAGGAGACCTGGAGTCT 3366
QY 250 eThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAl 270
Db 3367 GAGCGTGCTTCCAGGAATAAAGCTGAGAGACAGAAACGGGACCTTGGGAA-CAG----- 3420
QY 270 aPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSe 290
Db 3421 -----CTAGAGCGCTGAAACAGAGTTGGAGGACACGCTGGATTCCACAGCTGCCCA 3473
QY 290 rGlnAspTrpGlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysPr 310
Db 3474 GCAGGAGCTCAGTCAAAACGTGAGCAGGAGGTGAACATCTCTGAAGAGACCTGGAGGA 3533
QY 310 oLysMetAlaLysLeu-----CysGlnValLeuHisLeuSerAspAlaCysThrSerMe 328
Db 3534 GGAG---GCCAAGACCCAGGAGCCAGATCCAGGAGATGAGGAGAGCAAGCACTCACAGGC 3590
QY 328 tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIle-----TyrGluGl 345
Db 3591 CGTGGAGGAGCTGGCGGAGCAGCTGGAGCAGACGAAACCGGTGAAGCAAACTCGAGAA 3650
QY 345 uProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuVa 365
Db 3651 GCGAAGACGACTCTGGAGNACGAGCGGGGGAGCTGGCCACGAGTGAGTGCTGCT 3710
QY 365 l-----LysGlnIleLeuSerHisLysGluSerLeuVal----- 376
Db 3711 GCAGGGCAAAAGGGGACTCGGAGCACAAAGCGCAAGAAAGTGGAGCGCAGCTGCAGGAGCT 3770
QY 377 -----ArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
Db 3771 GCAGGTCAAGTTCACAGGAGGAGCGCGTGCAGAGCTGGCCGACAGAGTGCACCA 3830
QY 388 g-----ValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPhe-- 405
Db 3831 GCTGAGGTGGAGCTGGACAACTGACCGGGCTTCTCAGCCAGTCCGACGAGCAAGTCCAG 3890
QY 406 -----LeuGluHisThrGlnGlyLeuGl 413
Db 3891 CAAGCTCACCAAGGACTTCTCGCGCTGGAGTCCAGCTGCAGGACACTCAGGAGTGTCT 3950
QY 413 nPheAspCysPheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnGl 433
Db 3951 GCAGGAG-----GAGAACCGGAGAGAGCTGAGCCTGAGCACCACCAAGTCT----- 3993
QY 433 uSerLeuSerGlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSe 453
Db 3994 -----AAGCAGGTGGAGGACGAGAGAAATTCCTCCGGGAGCAGCTGGAGGA 4040
QY 453 rSerGluAspGluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrAr 473
Db 4041 GGAGGAGGAGGCCAAGCACAACTGGAGAGAGCAGATCGCCACCTCCATGCCCGAGTGGC 4100
QY 473 gAspIleGluLys-----ThrIleSerGlyIleLys 483
Db 4101 CGACATGAAAGAGAGAGTGGAGGACAGTGTGGGGTGCCTGGAAACTGCTGAGGAGTGAA 4160
QY 483 sLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysMetGl 503
Db 4161 GAGGAAGCTCCAGAGGAGCTGGAGGGCTTGAGCCAGCGGACGAGGAG----- 4212
QY 503 uLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAl 523
Db 4213 ----GTGGCGGCTAC---GACAGCTGGAGAGACCAAGAGC----- 4248
QY 523 aAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeu-- 542
Db 4249 -----CGGCTGCAGCAGGAGCTGGAGACCTTGCTGTGTGGGA 4283

QY 543 -LeuAspHis-----AsnTyrGluArgLysPheAspGl 553
Db 4284 CCTGGACCAACAGCGCGCTGCAACCTGGAGAGAGAGCAGAGAGTTTGACCA 4343
QY 553 uLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisAspValAlaLys 573
Db 4344 GTCTCTGGCGGAGGAGAGACCATCTCTGCCAAGTATGAGAGGAGCGGACCGGCTGA 4403
QY 573 sLysLysLeuAlaGluAspGlu 580
Db 4404 GCGGAGGCCCGAGAGAGAG 4425

RESULT 4

US-09-220-132-79

; Sequence 79, Application US/09220132

; Patent No. 6506607

; GENERAL INFORMATION:

; APPLICANT: Shyjan, Andrew W.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANC

; FILE REFERENCE: 07334-074001

; CURRENT APPLICATION NUMBER: US/09/220,132

; CURRENT FILING DATE: 1998-12-23

; PRIOR APPLICATION NUMBER: US 60/079,303

; PRIOR FILING DATE: 1998-03-25

; PRIOR APPLICATION NUMBER: US 60/068,821

; PRIOR FILING DATE: 1997-12-24

; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 79

; LENGTH: 5857

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-220-132-79

Alignment Scores:

Pred. No.:	1,31e-06	Length:	5857
Score:	154.00	Matches:	136
Percent Similarity:	39.64%	Conservative:	105
Best Local Similarity:	22.37%	Mismatches:	213
Query Match:	4.79%	Indels:	155
DB:	4	Gaps:	28

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-220-132-79 (1-5857)

QY 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 1840 GAAGAAACTCATCAGAGGAGATAAAGGCTCTGTATACCGCCACGGAAG----- 1890
QY 51 LeuAlaArgGluAsnMetArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 1891 CTTTCAAGAGAGACAGTCTATGAAAGCAAGCTGGAGCATGCCACGAGCGATGGAAGTCA 1950
QY 71 HisAlaLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
Db 1951 GATGTATAGTCTATGTAAGTCC-----AAACTG----- 1980
QY 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
Db 1981 -----GAGACTGCCATCGCATCCACGAGCGGCGATGGAAGTCA 2028
QY 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLysLeu----- 128
Db 2029 TCTTTTCAGCAAAAGGCTTGAACACGAGACGCGCAGATTTTGTGAATCTAAACACAAATA 2088
QY 129 ---CysLeuLeuLeuAlaLysHisAlaGlnGlyThrTyrSerSer-AspSerThrLe 147
Db 2089 GAGAAATGAGACTAGATTACCAACGAAATAGAAATTTGCAGATCAACAGACTCT 2148
QY 147 uPheGlyGluAspHisIleLysLeuSerAspGluSerProAsnIlePheTrpSer-- 166
Db 2149 GAACGGGCTGCCATCTGCTAAAGAGA-----TGAAGGC 2181

```
QY 167 -----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrProS 180
Db 2182 TTGAGGGCTAAACTGATGAAGTATTAAAGAAAGAAACAGTCTGGAAGCCATCAGG 2241
QY 180 exAspThrProGln-----ArgAsnArgLysArgValGlnTyrP 193
Db 2242 TCGAAACTGACAAAGCAGAGACACGATCTCGTAGAATGGAAGACACAGTATAACRAA 2301
QY 193 heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys 210
Db 2302 TTACAGGAAGCTGGAATAAGTAAGAGAGCTAGAGGTACTCGACAGCCAAATGCAATGAA 2361
QY 211 ArgLysLysAlaSerAspValThrAspPro---ArgValThrAspProValAsp 229
Db 2362 CAACCAAGGTTATTGATTAATTTTACATCACAGCTCAAGGCTACTGAAGAAAAGTCTTGG 2421
QY 230 AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu 245
Db 2422 GATCTTGATGCACTTCGGAAGCCAGTTCGGAAGGTAAATCGGAATGAAGAAACTTAGA 2481
QY 246 SerPro-----LysValIleThrLeuGln-----Ser 254
Db 2482 CAGCAGCTTGAGCGAGCTGAGAAACAGATTAAACATTTAGAGATTGAAAAGAAATGCTGAA 2541
QY 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
Db 2542 ACTAGCAAGGCTAGTAGCATTACCAGAGAGCTCCAGGGGAGAGAG--CTAAAGCTTACT 2598
QY 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly 294
Db 2599 AACCTTCAGGAAATTTGAGT-----GAAGT-CAGTCAAGTGAAGAGACTTTTGA 2648
QY 295 LysIleProAspGluSerGlnArgLeu-----HisThrValLeu 308
Db 2649 AAAAGAACTTCAGATTGTTGAAAGAAAAGTTGCTGAAGCTTCAGAGGAGCGACTCTCTGT 2708
QY 309 LysProLysMetAlaLysLeuCysGln-ValLeuHisLeuSerAspAlaCysThrSerMe 328
Db 2709 TCAGAGAAGTATGCAAGAACTGTAAATAGTTACACCAAGGAGGAAAGACAGTTTAACT 2768
QY 328 tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProIleTh 348
Db 2769 GCTGTCTTCTGACTTCGGAAGCTGAGAGAAACTTTAGCAGATATGAGGCGCAAAA----- 2823
QY 348 rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuValLysGlnI 368
Db 2824 -----TTTAGAGAGAA 2834
QY 368 eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
Db 2835 AGATGAGAGAGAAGACGACGCTGATAAGGCAAGGAAAGAACTG----- 2877
QY 388 gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi 408
Db 2878 -----GAAAAATGAC-----ATTGCAAGAAATAATGAAG----- 2904
QY 408 sThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerTh 428
Db 2905 -----ATGTCAGAGAGATAACTCTTCTCAG-----CTGAC 2933
QY 428 rLysLeuValAsnGlu-----SerLeuSerGlyAlaThrValArgAspGluLysIleAsn-- 446
Db 2934 AAAAATGAACGATGATTAAGTCTGAAAGAAAGAGAGATGTAGAAGAAATTACAGCTAAACT 2993
QY 447 -ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSe 466
Db 2994 TACAAGGCTAATGAAAATGCAAGT----- 3018
QY 466 rHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTy 486
Db 3019 -----TTTCTGCNAAGAAAGTATTGAGGACATGACTGCTCAAGC 3056
```

```
QY 486 rLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAs 506
Db 3057 TCAACAGAGCCAGCAAGAAAGCAGCTCTAAAAGAGCATGAGGAAGAAAGAAATTTGGAGAG 3116
QY 506 nMetTyrAlaAspLysLysGlnLysLeuGluThrSer-----LysSerValG1 522
Db 3117 GAAATTGCGGACCTGGAAAGAAATGGAACCAAGCCACCAACCACTGTCTCAGGAGCTGAA 3176
QY 522 uAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu----- 540
Db 3177 AGCCAGGTATGAGAGAGCCACT---TCTGAGACAAACCAAGCATGAGAAATCTCTACA 3233
QY 541 -----LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSe 556
Db 3234 GAACCTCCAGAGAGCGCTGCTGGAC-----ACAGAGGACCAAGCTCAAGGGCGCACGGGA 3287
QY 556 rGluLysAsnGluCysLeuLysSerLeuGluGlnMethIleAspValAlaLysLysLe 576
Db 3288 GGAGAACAGTGGCTTGTCTGAGAGCTGGAGAGCTGAGAAAGCTGAGAAAGCAAGCCGACAAACCCNA 3347
QY 576 uAlaGluAspGluAla 581
Db 3348 AGCTGCTCAACAGCG 3363
RESULT 5
US-08-875-435B-5
; Sequence 5, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; PRIOR FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6644
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-5
Alignment Scores:
Pred. No.: 4,25e-06 Length: 6644
Score: 150.50 Matches: 127
Percent Similarity: 35.02% Conservative: 94
Best Local Similarity: 20.13% Mismatches: 225
Query Match: 4.68% Indels: 187
DB: Gaps: 23
SEQ3_D1219_COPY_650_1270 (1-621) x US-08-875-435B-5 (1-6644)
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2995 CAGAGCTACAGCTCGAAAGAGTACCGCCGCGGCGCCAGATCAAGAGTTG----- 3045
QY 42 CysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 3046 -----GAGGACGACATCTCTGTCATGAGAGCATCAGAACCAACCAAGCTCTCAAAA----- 3093
QY 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
```



```
Query Match: 4.55% Indels: 162
DB: 4 Gaps: 27
SEQ3_D1219_COPY_650_1270 (1-621) x US-09-248-796A-6172 (1-1893)

QY 95 AspSerGly-ValSerPheGluGlnSerIleMet-----AspGlyValIleHi 110
DB 66 GATCTCTGGAAGTTGAAAAAGAGAGCTATGCTTAGTGAATAATCTTGAAACTGTCATTA 125
QY 110 eGluPheSerSerIleLeuSerSerIleGlyGlyGluGluAenGluValIleLeuCysLe 130
DB 126 GGAATACATGATGCAAAATTAAGGTTTAGAAGTAAGATCAACTCAATAAG----- 177
QY 130 uLeuLeuGluAlaLysHisAlaGln-----GlyThrTyrSerSerAspSerThrLeuPh 148
DB 178 -----GAGAAATCAATCAAAAGAGATAACAACACTCATAATGAACAAAGACTTCTTT 227
QY 148 eGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLe 168
DB 228 GAAACAAGATATTTGCAAAACTATCTCAAGATCAT-----GAACGGCGCAAACTCAATT 281
QY 168 uLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThrProGluArgAsnArgLy 188
DB 282 AGAAGTAAGGAGATCAATTTAGGAATTTGAAGCTAGTTTAGAGAAACACAAAT----- 336
QY 188 sArgValGlnTyrPheGluGluSerGluAlaSerProLysThrGlyAspGlyGlyAsnAl 208
DB 337 -----ACTGAGTCGCAACAAGCATTTGAAGAAAGAACAAAT-- 372
QY 208 aLysLysArgLysLysLeuAlaSerAspValThrAspProArgValThrAspProVa 228
DB 373 ----CAGATAAAGGAACCTCTCGAAACTATTAGTCATTAATAA---ACTGAATTAAGAC 425
QY 228 laePhePheGluArgLysLysLeuAlaSerGlyLysAspHis----- 240
DB 426 AAGTGTGTGATGCTTAAACAATCTCAAAAGGAATACAAAACCTTTAAAAACTAAAAATTC 485
QY 241 -----MetGlyAlaLeuGluSerProLysVal----- 249
DB 486 CGACACTGAACCTGAAATTTGAAAAACAATTTGGAAGAACTAGAAAAAGTGAAGCTGATTT 545
QY 250 -----IleThrLeuGlnSer-- 254
DB 546 ACAACACAGCAGATGAAAAATTTGAAAGGAATTCAGAAAGGGAATTTGCATTTGAAATCTGA 605
QY 255 -----SerCysLysSerSerGlyThrAspGlyThr----- 264
DB 606 ACTTGAAACAGTGAAATATAGTGGATTTGCTCAACACTTCTGAGCTTGTGCTGATTCACAAA 665
QY 265 -----LeuAspGlyAsnAspAl 270
DB 666 AACAGTGAATCATTTGAAAAAGAAAAAGGAAGAAATTTGCAATTTTATCGGCAACAATC 725
QY 270 aPheGlyLeuTyrSerMetGlySerHisLeuSerGlyIleProGluAspMetLeuAlaSe 290
DB 726 AAAGGAATTTGGAAGATTACATTTCAAAAGCACAGTCATGATTTTCGGAATAAACTAAAGCAT 785
QY 290 rGlnAspTrpGlyLysIlePro-----AspGluSerGlnArgArgLeuHisTh 306
DB 786 AACTGATGAGCTTAAAGAGAGAGCCAAACAGATTTGATTTGATTTCAAAAAGAAATTA---AC 842
QY 306 rValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAlaCysTh 326
DB 843 TGAGTTGGAAATATGATTTGACTTCCACTAAGAGGAATTTGGAACCTGAAAGACACAGAC 902
QY 326 rSerMetValGlyAsnPhe-----LeuGluTyrValIleGluAsnHisAr 341
DB 903 GAGTAAATTCAAAAACCTCGAGGAAAGGAAACAAAGAAATTTGTGAAACTCAATAAAGA 962
QY 341 gIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAl 361
DB 963 ATTAGAGTTATTGAAAAATGACACAAATAGTGGTCTAAAGGAAGAACTACTGGAAGAAAGTATC 1022
```

```
QY 361 aAlaLeu-----LeuValLysGlnIleLeuSerHisLysGluSerLeuVa 376
DB 1023 CAAGCTTGAGTCAGAAATTCAGATTTTATCCAAAGAGTTAGAGACAAATAATCGTCTCAT 1082
QY 376 lArgAlaAsnSerGluLeuAla-----PheLysCysSerArgVa 389
DB 1083 GAAACAACATGATGATGATTTGAAAGAACAAACAAAGGAAAAAATAATCAAGACTTCACAAAGT 1142
QY 389 lGluValAspTrpIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisTh 409
DB 1143 TACAAGGATTTACTTACTACCAAACTGAAA---TTAGATGAATTCGACAGAAATTAGA 1199
QY 409 rGlnGlyLeuGlnPhe---AspCysPheGlyThrAsnSer-----LysGlnSerVa 425
DB 1200 TGCTGCATTTGTCATTTCAAGATAAGTTTGAACACAGCTTCAGCCAAAGTTTAAACAATCTAC 1259
QY 425 lValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysI 445
DB 1260 TTCTGATTTGGAAGCAGCAAAATTAAGAAATTTGAATATTTTGTATTAGTGAAGAAAAAACA 1319
QY 445 eAsn-----ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGl 462
DB 1320 TGAACAAGAACTTGAATAAACTTACAAACAACACGCTGAACTTGAAAAAACTGATTGATGA 1379
QY 462 uLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyI 482
DB 1380 TAAAAATGCT-----GATTCACAGTAAGACAAATTTCTGATCT 1415
QY 482 eLysLysLysTyrLysLys-----GlnValGlnLysLeuValGlnGl 496
DB 1416 TGAATAAAGCCAAACAAGAAAGTATTTGACCTTGAACGCAACTTGAAAAAACTTACCACAGA 1475
QY 496 uHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGl 516
DB 1476 ACTTGACTACTAATAACTGGAATTA-----CAAAAACATGAAAAAATG-- 1518
QY 516 uThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGl 536
DB 1519 -----AACAAATGA 1526
QY 536 nValGlyAspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSe 556
DB 1527 ATTCACCAATTTGAGTAAATCATCATGAGAAATTAATACTAAATATAGACAAATTTGAAAA 1586
QY 556 rGluLysAsnGluCysLeuLysSerLeuGluGlnMet-----HisAspValAlaLys 573
DB 1587 TGAAGAAGATGATCTTGTCAATCATCATGAAAAACTCAATGTTGAACATAAAGAAACGTC 1646
QY 573 sLysLysLysLeu-----AlaGluAspGlu-----AlaCysTrpIleAsnAr 586
DB 1647 AACAAAAATTACAAAAATGTTGAAGAAAGAAAAATTAAGAAAAATCACAGCAAAATATGAACTCGA 1706
QY 586 gIleLysSerTrpAlaAlaLysLeuLysValCysValPro 599
DB 1707 ACTCAAGAAATTAACAAGATAAACTTTAAGAAATTTGTGTTCTT 1746

RESULT 7
US-09-949-016-1054
; Sequence 1054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```


QY 559 AenGluCyLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGlu 578
: : : : :
Db 4735 AGTGAACACTACAGACTTCCCTAGAGAA-----GCAGAGGCATCTCTTGACAT 4782
: : : : :
QY 579 AspGluAla-----CysTrpIleAsnArgIleLysSer 589
: : : : :
Db 4783 GAAGAAGGCAAAATTTCTCGCATTCAACTTGAGCTAAATCAGGTGAATCT 4833
: : : : :
RESULT 8
US-09-949-016-2264
; Sequence 2264, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2264
; LENGTH: 6016
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2264

Alignment Scores:

Pred. No.: 1,22e-05 Length: 6016
Score: 146.00 Matches: 136
Percent Similarity: 35.60% Conservative: 105
Best Local Similarity: 20.09% Mismatches: 256
Query Match: 4.54% Indels: 180
DB: Gaps: 27

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-949-016-2264 (1-6016)

QY 19 LysHisValGluLysIleLysLeuGluSerCysSerGluArg---ThrLysIlePheArg 37
: : : : :
Db 3025 AAACATGCCACAGAACAAAGGTGAAACCTCACAGAGAGATGGCAGGTCTGGATGAA 3084
: : : : :
QY 38 LeuTyrSerValCysThrValGluGluLysAlaLeuLysAlaLeuAlaArgGlnAsnMetArg 57
: : : : :
Db 3085 ACCATTGCTAAGCTGACCAAGGAGAGAAAGGCTCTCCAGGAGGCCACCAGCAGACCCCTG 3144
: : : : :
QY 58 GlnAsnLysAlaValGluAlaLeuAsnArgSerLeuThrHisAlaLeuMetTrpGly 77
: : : : :
Db 3145 GATGACCTGCAGATGGAGAG-----3165
: : : : :
QY 78 AlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGly 97
: : : : :
Db 3166 -----GACAAAGTCAACACCTGACCAAAAGCTAAAAACC-----3198
: : : : :
QY 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
: : : : :
Db 3199 ----NAGCTAGAACAGAGTGGACGAT-----3222
: : : : :
QY 118 SerLysGlyGlyGluGlnAsnGluValLysLeuCysLeuLeuGlu-----AlaLys 135
: : : : :
Db 3223 CTGAGAGGATCTCTGAAACAGAAAGAAACTTTGATGACTTAGAAGAGCCAGAGA 3282
: : : : :
QY 136 HisAlaGlnGly-----ThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHis 152
: : : : :
Db 3283 AAACCTGGAGGGTGACCTAAATTTGGCCCAAGAAATCCACAATGGATGATACAGAAATGACAAA 3342
: : : : :
QY 153 IleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuGlyGlyLys 172
: : : : :
Db 4318 GATGCAGAGAACATGTAGAGCTGTGAATTCCTCAATATGTGCTTCTTTGAAAAGACAAAG 4377
: : : : :
Db 4377

Db 3343 CAGCAACTTAATGAGAAACTCAAAAGAAAGAGATTTGAAATGAGCAATCTGCAAGGCCAAG 3402
: : : : :
QY 173 AsnProMetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGln--- 191
: : : : :
Db 3403 ATTGAAGATGAACAGCCCTTTCATACAGCTACAAAAGAGATCAAAAGAAATTACAGGCC 3462
: : : : :
QY 192 TyrPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArg 211
: : : : :
Db 3463 CGCATTGAGGAGCTGGAGGAGAAATCGAGCAGCGCGGCTCCCGGGCCCAAGACAGAG 3522
: : : : :
QY 212 LysLysAlaSerAsp-----AspValThrAsp-----220
: : : : :
Db 3523 AAGCAGCGCTCTGACCTCTCCGGGAGCTGGAGGAGATCAGTGAGAGGCTGGAAGAGCC 3582
: : : : :
QY 221 ProArgValThrAspProValAsp---AspAspGluArgLysAlaSerGlyLysAsp 239
: : : : :
Db 3583 GGTGGGGCCACTTCAGCCCGAGATTGAGATGAACAGAGCGGAGGCTGAGTTCCAGAAA 3642
: : : : :
QY 240 HisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSer 259
: : : : :
Db 3643 ATGCGCAGGGACCTGGA-----GAGTCCACCTCGCAGCAGCAA 3681
: : : : :
QY 260 GlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHis 279
: : : : :
Db 3682 GCCACGGCAGCTGCTCTTCGGAAGAGCAGCAGATAGTGTGGCTGAGCTTGGGAGCAG 3741
: : : : :
QY 280 IleSerGlyIlePro-----284
: : : : :
Db 3742 ATTGACGCTTCAGCGGCTCAAGCAGAGCTGGAGAGAAAGAGAGTGAGTGAGATG 3801
: : : : :
QY 284 -----284
: : : : :
Db 3802 GAGATCAATGACCTTGCTAGTAACATGGAGACTGTCTCCAAAGCCAAGCAAACTTTGAG 3861
: : : : :
QY 285 -----GluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAsp 298
: : : : :
Db 3862 AAATGTCCGCGCACCTAGAGCCAGCTTAGTGAATA-----AAACAAGGAA 3912
: : : : :
QY 299 GluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu-----315
: : : : :
Db 3913 GAAGAGCAACACGCTTAATAATGAGTTGTCCAGCCAGAGGACGCTTTACACACAGAA 3972
: : : : :
QY 316 -----CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsn 331
: : : : :
Db 3973 TCAGGTGAGTTTTCACGACGACTAGTGAAGAAAGATGCTATGTTCTTCAGCTATCCGA 4032
: : : : :
QY 332 -----PheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu-----345
: : : : :
Db 4033 GCACAAACAGCAATTTACACACAGATTGAAGAAATTAAGAGGCGACTAGAGAGGAGACT 4092
: : : : :
QY 346 -----ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeu 363
: : : : :
Db 4093 AAGGCCAAGAGCAGCTCTGCCCATCTGCCCTGCAGTCAGCC---CGCCATGACTGTGACCTG 4149
: : : : :
QY 364 LeuValLysGlnIleLeuSerHisLysGluSer-----LeuVal 376
: : : : :
Db 4150 CTGCGGAGACAGTATGAGGAGAGCAGGAAGCAAGGCTGAGCTGCAGAGGGGAATGCC 4209
: : : : :
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSer 396
: : : : :
Db 4210 AAGGCCAAGCAGTGAAGTTGCCCGAGGAGGAGCAAGTACGAGAGCGGAGCCCATC-----4263
: : : : :
QY 397 IleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCys 416
: : : : :
Db 4264 -----CAGCCACACAGAGGAGCTGGAG-----4284
: : : : :
QY 417 PheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSer 436
: : : : :
Db 4285 -----GAGGCCAAGAAAGACGTAGCCAGCCAGCGCTCTGCGAG 4317
: : : : :
QY 437 GlyAlaThrValArgAspGluLysIleAsnThrLys-----Ser 449
: : : : :
Db 4318 GATGCAGAGAACATGTAGAGCTGTGAATTCCTCAATATGTGCTTCTTTGAAAAGACAAAG 4377
: : : : :
Db 4377

QY 450 MetArgAsnSerSerGluAspGluGluCysMet-----ThrGluLysArg 464
DB 4378 CAGAGGCTCAGAAATGAGTGAAGGAGCTCATGATGATGGAACGATCTAATGTGCTGCC 4437
QY 465 CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLys 484
DB 4438 TGCATAGCTCTCGATGAAGAAAGAAACTTTGACAAGGTTCTGGCAGAAATGGAACAG 4497
QY 485 LysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLys----- 501
DB 4498 AAGTATGAGGAAACTCAGGCTGAACCTTGAGGCTCCAGAGGAGTGGCGTTCTCTCAGC 4557
QY 502 -----MetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSer 518
DB 4558 ACTGAGCTGTTCAGGTGAAGATGCTACGAGGAATCCCTGGCATCTCTGAACTCTA 4617
QY 519 LysSerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
DB 4618 AAGCGA---GAGAATGAAGAACTTACACAGAGGATTTCTGACCTGACAGAGCAATTGCA 4674
QY 539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys 558
DB 4675 GAGGTTGGAAGCATATCCATGAACTGGAGAAAGTAAAGAAACAATCTGATCATGAGAAG 4734
QY 559 AsnGluCysLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGlu 578
DB 4735 AGTGAACACTACAGACTTCCCTAGAGAA-----GCAGAGGCATCTCTTGAGCAT 4782
QY 579 AspGluAla-----CysTrpIleAsnArgIleLysSer 589
DB 4783 GAAGAAGGCAAAATCTTCGATTCACACTTGAGCTAAATCAGGTGAATCT 4833

RESULT 9

US-09-248-796A-183
; Sequence 183, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 183
; LENGTH: 4947
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-183

Alignment Scores:
Pred. No.: 1,31e-05 Length: 4947
Score: 144.50 Matches: 132
Percent Similarity: 36.72% Conservative: 103
Best Local Similarity: 20.62% Mismatches: 243
Query Match: 4.49% Indels: 163
DB: 4 Gaps: 28

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-248-796A-183 (1-4947)

QY 60 LysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSer 79
DB 2698 AAAAAATTGCAACGATTGATTAATCATTCAGAGAGAAATTCGAATTTGTGGCAACACTC 2757
QY 80 TyrLeuPhe-----AspLysLeuAspHisPheHisSerSer 91
DB 2758 TTTTATTTTATTCAAACCAATCCATTTATTGACAAAATTTG-----TTTAATTCAT 2811

QY 92 GluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHisGlu 111
DB 2812 CCATATACTAAAAAATCAGCAAAAATCTGTCAAGATTTCGTCCAGTCTGTGCTCAATTA 2871
QY 112 Phe-----SerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLys 127
DB 2872 TTCCCAAGTAAGAGATTTCATCAATTACTTATCATCTTCGTGAAGAGTAGTACITTTTGGTCAA 2931
QY 128 LeuCysLeu---LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThr 146
DB 2932 CTTATCATCCAGTTGATGCAAAAATGATACAGCTAACTCAACAATTTGGGTGACATCACC 2991
QY 146 ----- 146
DB 2992 AAGTTGCATTTGACAACTGGATAGATTTCTTTACTAATTTTCAATATCATCATCTCCAG 3051
QY 147 -----LeuPheGlyGluAspHisIleLysLeuSerAsp----- 157
DB 3052 AGACAACACTTGAAGACACTTTTGGCAAGTTTGTATAGTATTGTTGATATGATATGAACAA 3111
QY 158 -----GluGluSerProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
DB 3112 GTGGAATTTTGAATCAGATCCAAATCAGAATCTCAATCAAAATCATT-----GATCATGAG 3165
QY 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
DB 3166 ATGAAAGTTTGTGCTAGTGAAGAGTCTCGTGATATTTTCCACACAGGAGCTATTTCAG 3225
QY 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysAla 214
DB 3226 TTGCCAGAAAGTTAGTAAATAGTTT-----GTGGGCATTTGATGTCTTAAGGGAGACT 3279
QY 215 SerAspAspValThrAspProArgValThrAspProValAspAspGluArgLys 234
DB 3280 TGCAGTGATTTGCTTTCATGTTA-----CAAAAAAT 3312
QY 235 AlaSerGlyLys-----AspHisMet----- 241
DB 3313 GCTTCTGGAATAAAGACCTTTGTGCAAAATCCCTGACCATGTGAAATTTGATTTGCTGCA 3372
QY 242 GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThr 261
DB 3373 GGTATTATGTGCTCAAGAAGTTCCCGCAAAATCTGATCAACAACATTTAGCGGTT 3432
QY 262 AspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSer 281
DB 3433 GCTGGGGTTATA-----TTTGTGAACACTATCTCTGATCT---ATTTCG 3474
QY 282 GlyIleProGluAsp-----MetLeuAlaSerGlnAspTrpGlyLysIleProAsp 298
DB 3475 CAAGTACCAGAGAACTATGGAATTTTGAAGTGGTAAACAATGATACCCAGAAAGCTAAAGC 3534
QY 299 ---GluSerGlnArgGluHisThrVal----- 307
DB 3535 AAAGACAATTTGCGATACTTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3594
QY 308 -----LeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAsp 323
DB 3595 AATGACAATTTTGAACCACTTGAATGAG-----TACATTATGCAAGTACCGAC 3645
QY 324 AlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyr 343
DB 3646 ACCGTTAAGTCAATCATTAGTCAGGCAATTTAATGTGGTGAATTTGAAACCGTGTAT 3705
QY 344 Glu-----GluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrp 359
DB 3706 GAATTACATGATTATGATGATCTGTGACTCATCAGAGACCTTAAGTTGACTATCAGTGTA 3765
QY 360 IleAlaAlaLeu---LeuValLysGlnIleLeuSerHisLysGluSerLeuValAlaGala 378
DB 3766 AACTCATTCATTCATTAAGAAATCCATCTTGCATAAATGTCGATATTATTACACAGGC 3825
QY 379 AsnSerGluAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398

```
Db 3826 AATGATGATCAATTTGACAAA---ACATCGTCGAGGTGCAAAAATTTG-----TTGATT 3876
Qy 399 SerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGly 418
Db 3877 TCCCAACAAGACATGTTGACATGACCGATTGTCATCGGTTACTTGTGAACCTTAATCTC 3936
Qy 419 ThrAsnSerGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAla 438
Db 3937 ACCACACAGAGGAGCTATAGTTGATTCAAA----- 3969
Qy 439 ThrValArgAspGluLysLeuAsnThrLysSerMetArgAsnSerSerGluAspGluGlu 458
Db 3970 -----ACTAAACATTTGTTCCACCAAGCC----- 3993
Qy 459 CysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp----- 474
Db 3994 -----AAACGCTGCTTGTGTATATCATTTAGAGTCGAGGAGACAGACAGT 4041
Qy 475 -----IleGluLysThrIleSerGlyIleLysLysLysLysGlnValGln 491
Db 4042 GATGATTTATTGGAATTTAATATATCGGCATCAACCAAGCCACGAAACAAAGATTCAAG 4101
Qy 492 LysLeuValGlnGluHisGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLys 511
Db 4102 GAGATTGTTCAATATGAAAA-----GCGGAGCAA 4131
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAlaValIle-----ArgIle 528
Db 4132 GATATTTCCCTCAACACAGCAGAGTGTTACTGCTAATCTGCTAATAAAGAAATCT 4191
Qy 529 ThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGlu 548
Db 4192 ACCAGACCATATTCGGGTACATCATTAAGGGGACTTGTGCAATTTAACTATCAT----- 4245
Qy 549 ArgLysPheAspGluLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMet 568
Db 4246 -----GAATTGAAAAAATGTTGTTGGAGATTATTCTTAAAGCTAGAAATCAATG 4293
Qy 569 HisAspValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLys 588
Db 4294 GGGGAGTTGACTGCTGTAATAACTCATTTCAA-----ACATGCTCAATCAATAGCT 4344
Qy 589 SerTrpAlaAlaLysLysValCysValProIleGlnSerGlyAsnAsnLysHisPhe 608
Db 4345 A--TGACATTAACAAAGAGTCTCAAAGACAATGCGGTGGCAACAATTTGCAAGTTT 4402
```

RESULT 10

```
US-08-353-700-2
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TRANSCRIPTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10136 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-2

Alignment Scores:
Pred. No.: 4,628-05 Length: 10136
Score: 144.50 Matches: 150
Percent Similarity: 35.20% Conservative: 101
Best Local Similarity: 21.04% Mismatches: 237
Query Match: 4.49% Indels: 225
DB: 1 Gaps: 37

SEQ3_D1219_COPY_650_1270 (1-621) x US-08-353-700-2 (1-10136)

Qy 7 LeuPheGlySerSerLeu-----AsnProSerHisAsp 17
Db 5034 GTATTTTCGACGAGTCTGCAGGAGGAGAAATCTGACAGGAAAGAAACCCCTTCGCGCCCA 5093
Qy 18 ValLysHisValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 37
Db 5094 GCGAAGGTGTGAA-----GAGCTTGAGTCCCTCTGTGAG-----GTGTACCGG 5138
Qy 38 LeuTyrSerValCysThrValGluLysAlaLeuLysLysLysLysLysLysLysLysLys 57
Db 5139 ---CAGTCCCTCGAAGCTAGAGAGAA-----ATGGAAGTCAAGGATTATG 5186
Qy 58 GlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGly 77
Db 5187 AAAAAAAGGAAATTCAGAGCTCGAGCAG-----TTATTAAGTTCT 5228
Qy 78 AlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGly 97
Db 5229 GAAAGCGAAGAGCTTGCTGCTTTAGGAAGCAGATTTGTCAGAAAATGAACAG----- 5282
Qy 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
Db 5283 -----TGCAACACAGAGCTGCAAGCGTGAGTCTGGAGATGGAGTCCAAGTTGGCG 5333
Qy 118 SerLysGlyGlyGluGluAsnGluValLysLysLysLysLysLysLysLysLysLysLys 137
Db 5334 GCAGAAAAAGAAACAGACGGAACACTGTCACTGTGAGCTGGAAGTGAAGTACGACGCTCA 5393
Qy 138 GlnGlyThrTyrSerSerAspSerThrLeuPheGly-----GluAspHisIleLys 154
Db 5394 CAAGGTCTGGACTTAAGTTCTCGGTCTTTTGTGGCATCGACACAGAGATGCTATTCA 5453
Qy 155 LeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuLeuGlyLysAsnPro 174
Db 5454 ---GCCCAAAATGAGAGCTGTGACATA-----TCAAAA----- 5483
Qy 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArg---ValGlnTyrPhe 193
Db 5484 -----GAACATCTCTCAGAACTACAGAAAGAACACCAAGCATGATGTTTCATCAGATT 5537
Qy 194 GluGlySerGluAlaSer-----ProLysThrGlyAspGlyGly 206
Db 5538 TGTGATAAGATGCTCAGCAGACCTCAATCTAGACATTGAGAAATAACTGAGACTGCT 5597

QY 207 AsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgValThrAspPro 226
Db 5598 GCAGTGAACCCACGAGAGAGTCTCTGGGGAACAGTCCCCAGATACCAATTATGAGCCT 5657
QY 227 ProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer 246
Db 5658 CCAGGGGAAGAT-----AAACCCAGGGC----- 5681
QY 247 ProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAsp 266
Db 5682 -----TCTTCAGATGCAATTCT-----GAATTGTCAATTTCT 5714
QY 267 GlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisLysSerGlyLysProGluAsp 286
Db 5715 GGTCCCTAATGCTTTG-----GTACCTATGAT 5741
QY 287 MetLeuAlaSerGlnAsp-----TrpGlyLysIleProAsp 298
Db 5742 TTCTTGGGGAATCAGGAAGATATCCATAATCTTCAACTGCGGTAAAGAGACATCAAT 5801
QY 299 GluSerGlnArgLeuHisThrVal-----LeuLys 309
Db 5802 GAGAATTGATTAATCTTCAATGATAGAGGCGGTGACAGAAAGTTGAAAGTTGCTA 5861
QY 310 ProLysMetAlaLysLysGlnValLeuHisLeuSer----- 322
Db 5862 AATGAATGAAGAAATAGACTCAAACTCCATTTACAGGAGGTACAACTAATGACCAAA 5921
QY 323 ---AspAlaCysThrSerMet-----ValGlyAsnPhe----- 332
Db 5922 ATTGAAGCATGCATAGAAATTTGGGGAATTTAGTTGGGGAATTTAGAAAGAAACTCAGAT 5981
QY 333 -----LeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaThr 348
Db 5982 TTAAGTGAATAATTTGAAATATTTTCTTGATCACCAGGAGTACTCCAGAGATAGAA 6041
QY 349 ThrPheGlnAlaPheGlnIleAlaLeuSerTrp----- 359
Db 6042 ACTTCTGAAGGCTCAATCTGATTTAGAAATGCATGCAGATAATCATCAGCTGAAGAT 6101
QY 360 IleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
Db 6102 ATTGGAGATATATGTCACAGGTGAATGCAGCTGGAAGGAGAGATTTCTTGATGGAA 6161
QY 380 SerGluLeuAla---PheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398
Db 6162 AATGAGCTGATGATGATCAGATCGGAGAAAGTAGCATTCAGCAT----- 6206
QY 399 SerCysMetLysSerLeuPheLeuGlu-----HisThrGlnGlyLeu 412
Db 6207 -----GAAGCCCTTACTCTGGAGGCTGACTTAGAGGTAGTTTCAACAGAGAAGCTA 6257
QY 413 GlnPheAspCysPheGlyThrAsnSerLysGlnSerValSerValSerThrLysLeuValAsn 432
Db 6258 TGTTTAGAAAAGACAATGAATAATAGCAGAAGTTATTGTC-----TGCCTTGA 6308
QY 433 GluSerLeuSerGlyAlaThrValArg----- 441
Db 6309 GAAGAACTCTCAGTGTCTCAGTGAAGAACCCAGCTTCTGCGAGAAATTAGATATATG 6368
QY 442 -----AspGluLysIleAsnThrLysSerMetArg 451
Db 6369 TCAAAAAAACCCAGGCTGATCATCTTCTGAAAAAATGAAGGAGAAACACACAGAG 6428
QY 452 AsnSerSerGluAspGluGluCysMet----- 460
Db 6429 CTTGAGTCTCATCAAGTGAAGTCTCTCATTCATTGATTCAGGTGCGAGAGGAGGTAAG 6488
QY 461 -----ThrGlu-----LysArgCys 465
Db 6489 GAAAGACGGAATCTCTTCAAGATTTGTCCTCTGATGTGATGAGTGTGTTAAAGACAAA 6548

QY 466 SerHisTyrSerThrAlaThrArgAspIleGluLys-----ThrIleSerGlyIle 482
Db 6549 ACTCATCTCCAGGAAAGAGTGCAGAGTTTGGAAAGAGACTCACAGGCACTGTCTTTGACA 6608
QY 483 LysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGlu-----GluLys 500
Db 6609 AATGTGAGCTGGAACCAATTTGCACAACTGAATAAGAGAAAGAAATGCTTGTTCAG 6668
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLysLys-----GlnLysLeuGluThrSer 518
Db 6669 GAATCTGAAAGCCTGCAGCCAGACTGAGTGAATCAGATTATGAAAGCTGAATGTCTCC 6728
QY 519 LysSerValGluAlaValIleArg-----IleThrCysSerArgThr 533
Db 6729 AAGCCCTTGGAGCCGCTGCTGGAGAAAGGTGAGTTCGCATTGAGGCTGAGCTCAACA 6788
QY 534 SerThrGlnValGlyAspLeuLysLeuAspHisAsnTyrGluArgLysPheAspGlu 553
Db 6789 CAGGAGGAAGTG-----CATCAGCTGAGAAGAGCGCATCGAGAAA 6827
QY 554 IleLys-----SerGluLysAsn-----GluCysLeuLys 563
Db 6828 CTGAGAGTTTGGCATTGAGGCCGATGAAAGAGCAGCTGCACATCGCAGAGAACTGAAA 6887
QY 564 SerLeuGluGlnMethHisAspValAlaLysLysLysLeu 576
Db 6888 GAACGCGAGCGGAGAAATGATTCATCTTAAGGATAAAGTT 6926
RESULT 11
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: PCT/US95/16216
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-2

Qy	299	GlusSerGlnArgArgLeuHisThrVal	LeuLys	309
Db	5802	GAGAAATTGGATTTACTTCATGTGATGATGACGACCGTGACAGAAAGTTGAAAGTTTGCTA	5861	
Qy	310	ProLysMetAlaLysLeuCysGlnValLeuHisLeuSer		322
Db	5862	AATGAATGAAGAAATTAGACTCANAACCTCCATTTACAGGAGGTACAACTAATGACCCAAA	5921	
Qy	323	---AspAlaCysThrSerMet	ValGlyAsnPhe	332
Db	5922	ATTGAAGACTGCATAGAAATTGGAAAAATAATAGTTGGGAACTTAAAGAAGAAAACTCAGAT	5981	
Qy	333	-----LeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaThr		348
Db	5982	TTAAGTCAAAAATTTGGAAATATTTCTTGTGTGATCACCAGGAGGTACTCCAGAGAGTAGAA	6041	
Qy	349	ThrPheGlnAlaPheGlnIleAlaLeuSerTrp		359
Db	6042	ACTTCTGAAGCCCTCAATTCCTGATTTAGAAATGCGATGCAGATAAATCATCACGCTGAAGAT	6101	
Qy	360	IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn		379
Db	6102	ATTGGAGATAATGTGGCCAAAGGTCAATGCAGCTGGGAAGAGAGAGATTCTTGTGATGTGGAA	6161	
Qy	380	SerGluLeuAla---PheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu		398
Db	6162	ANTGAGCTGAGTAGGATCAGATCGGAGAAAGCTAGCATTCAGCAT		6206
Qy	399	SerCysMetLysSerLeuPheLeuGlu	HisThrGlnGlyLeu	412
Db	6207	-----GAAGCCCTCTACCTGGAGGCTGACTTAGAGCTAGTTCAAACAGAGAAGCTA	6257	
Qy	413	GlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsn		432
Db	6258	TGTTTAGAAAAAGACAAATGAATAATAGCAGAGAGGTATTGTC		6308
Qy	433	GluSerLeuSerGlyAlaThrValArg		441
Db	6309	GAAGAACTCTCAGTGGTCACAGTAGAGAGAAACCAGCTTCGTGGAGAATTTAGATACTATG	6368	
Qy	442	-----AspGluLysIleAsnThrLysSerMetArg		451
Db	6369	TCAAAAAAAACCCAGGCACCTGGATCATGTTGCTCTGAAAAAATGAAGGAGAAAAACACAAGAG	6428	
Qy	452	AsnSerSerGluAspGluGluCysMet		460
Db	6429	CTTGAGTCTCATCAAGTGAAGTGTCTCCATTGCATTTCAGGTGGCAGAGGCAGAGGTGAAG	6488	
Qy	461	-----ThrGlu	LysArgCys	465
Db	6489	GAAMAGACGGAACTCCTTCAGACTTTGTCTCTCATGTGAGTGAGCTGTTTAAAAAGACAAA	6548	
Qy	466	SerHisTyrSerThrAlaThrArgAspIleGluLys	ThrIleSerGlyIle	482
Db	6549	ACTCATCTCCAGAAAGAGCTGCAGAGTTTGGAAAAAGGACTCACAGGCACCTGTCTTTGACA	6608	
Qy	483	LysLysLysTyrLysLysGlnValGlnLysLeuValGlnHisGlu	GluLys	500
Db	6609	AAATGTGAGCTGGAAAAACCAATTGCACAACTCGAATAAAGAGAGAAGAAATTTGTTGTCAG	6668	
Qy	501	LysMetGluLeuLeuAsnMetTyrAlaAspLysLys	GlnLysLeuGluThrSer	518
Db	6669	GAATCTGAAAGCCCTGCAGGCCACAGCTGAGTGAATTCAGATTATGAAAAGCTGGAATGTCTCC	6728	
Qy	519	LysSerValGluAlaAlaValIleArg	IleThrCysSerArgThr	533
Db	6729	AAGCCCTTGAGGCCCGCACCTGGTGGAAGAAAGGTGAGTTCCCATTTGAGCTGAGCTCAACA	6788	
Qy	534	SerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGlu		553
Db	6789	CAGGAGGAAGTG-----CATCAGCTGAGAAAGAGGCATCGAGAA	6827	

```
QY 554 lleyys-----SerGluLysAsn-----GluCysLeuLys 563
Db 6828 CTGAGAGTCCGATTCAGGCGCGATGAAAGAACAGCAGTCACATCGCAGAGAAACTGAAA 6887
QY 564 SerLeuGluGlnMethHisAspValAlaLysLeuLysLeu 576
Db 6888 GAACGGAGCGGAGAGATGATTCACCTAAGGATAAGATT 6926

RESULT 12
US-09-917-254-40
; Sequence 40, Application US/09917254
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7224(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/222,093,
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 5574
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-917-254-40

Alignment Scores:
Pred. No.: 1,85e-05 Length: 5574
Score: 144.00 Matches: 133
Percent Similarity: 35.38% Conservative: 92
Best Local Similarity: 20.91% Mismatches: 216
Query Match: 4.48% Indels: 197
DB: Gaps: 28

SEQ3_D1219_COPY_650_1270 (1-621) x US-09-917-254-40 (1-5574)
QY 22 GluLysIleLysIleGluSerCysSerGluAUGThrLysIlePheArgLeuTyrSerVal 41
Db 2554 CAGAACTGCACTTGGAGGAGTTCAGGCTGAGGCGGAGGAGTCAAGAACTG-----2604
QY 42 CysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 2605 -----GAGGATGAGATCCTGGTCATGGATGATCAGAACTAATAACTATCAAAA---2652
QY 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db 2652 -----2652
QY 82 PheAspLysLeuAspHisPheHisSerSerGluThrThrProAspSerGlyValSerPheGlu 101
Db 2653 -----GAA 2655
QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 121
Db 2656 CGAAAACTCTTGAGGAGAGGATTAGTACTTAACAGCAAAATCTTCAGAA-----2706
QY 122 GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
Db 2707 GAGGAAGAAAGGCCAAGAACTTACCAAGCTGAAACAAAGCAT-----2751
QY 142 SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerPro 161
Db 2752 -----GAATCTATGATTTCAAGACTGCGGCTAAGAGGAGAGAGAGAGC 2802
QY 162 AsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAsp 181
Db 2803 CGACAGAGCTGAGAGAGCTG-----AAACGGAAGCTGGAGGAGTATCCACGAC 2853
QY 182 ThrProGlnArgAsnArg-----LysArgValGlnTyrPheGluGly 195
Db -----2874
```

```
Db 2854 TT-CCACGAGCAGATCGCTGACCTCCAGGCGGAGATCGCAGAGCTCAAGATGACAGCTGGC 2912
QY 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSer 215
Db 2913 CAAGAGGAGGAGGAGCTGCAGGCGCCCTGGCCAGGCTTGACGATGAAATCGCTCAGAA 2972
QY 216 AspAspVal-----ThrAspProArgValThrAspProVal-Asp-----229
Db 2973 GAACAATGCCCTGAAGAAGATCGGAGAGCTGGAGGCGCCACATCTCAGACCTCCAGGAGGA 3032
QY 230 ---AspAspGluArgLysAlaSerGlyLys-----As 239
Db 3033 CTGAGCTCAGAGCGGCGCCAGGAACAAGCGCTGAAACAGCAGAGGAGACCTCGCGCA 3092
QY 239 phisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSe 259
Db 3093 GGAGCTGGAGGCGCTTAAGACAGAGAGCTGGAAGACACACTGGACAGCACACCCAC-TCAGC 3151
QY 259 rGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHi 279
Db 3152 AGGAGCTCAGGCG-----CAAGAGGAGGAGCAGG 3178
QY 279 sileSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIlePro-AspG 299
Db 3179 AGGTGACGTGCTGAAGAAGGC-----CCTGATG 3208
QY 299 luSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnVal 319
Db 3209 AAGAGAGCGGCTCCCATGAGGCTCAGGTCAGGAGATGAGGAGGAGGAGGAGGAGGAGG 3268
QY 319 euHisLeuSerAspAlaCysThrSerMetValGlyAsnPhe-----332
Db 3269 TG-----GAGGAGCTCAGACAGAGCAGCTTGAGCAGTTCAAGAGGCGCAAGCGCAACC 3319
QY 333 -----LeuGluTyrValIleGluHisArgIleTyrGluGluP 346
Db 3320 TAGACAAGAAATAAGCAGACGCTGGAGAA-----GAGAACGCGACAGCTGGCGGGAGC 3373
QY 346 roAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuVal 366
Db 3374 TCGCGGTCTCGGCGCAGGC-----A 3394
QY 366 yGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn---SerGluLeuAlaPheL 385
Db 3395 AGCAGAGGCTGGAACATAGAAAGAGAGCTGGAGCGCAGGTCGAGGAGCTGCAGTCCA 3454
QY 385 yCysSer-----ArgValGluVal---AspTyrIleTyrSerIleLys 399
Db 3455 AGTGACAGCATGGGAGCGGCGCGGCGGAGCTCAATGACAAAGTCCACAGCTGCAGA 3514
QY 399 exCysMetLysSerLeu-----PheLeuGluHisThrGlnGly-----LeuG 413
Db 3515 ATGAAGTTGAGAGCTGACAGGAGTCTTAACGAGCGCGGAGGAGGAGGAGGAGGAGGAGG 3574
QY 413 lnPheAspCysPheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnG 433
Db 3575 CCAAGGACGCTGGCGTCCCTCAGTCTCCAGCTCCAGGACACCAGGAGGAGTGTCTCAAGAAG 3634
QY 433 luSerLeuSerGlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnS 453
Db 3635 AAACCGCGCAGAAAGTCAACGTGTCTACGAAGCTGCGCAGCTGGAGGAGGAGGAGGAGGAG 3694
QY 453 exSerGluAsp-----GluGluCysMetThrGluLysArgCys-----SerHisTyrS 469
Db 3695 GCTTGAAGACCAAGCTGGAGGAGGAGTGGAGGCGCAAGCAGACACCTGGAGGAGGAGGAGG 3754
QY 469 exThrAlaThr-----ArgAspIleGluLysThrI 479
Db 3755 CCACCTCTCAACATCCAGCTCTCCGACTCGAAGAAGAGCTGCAGGAGCTTTGCGAGCACCG 3814
QY 479 leSerGlyIle-----LysLysLysTyrLysLysGlnValGlnLysLeuValGlnG 496
Db 3815 TGGAAAGCTTGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3874
```

```
QY 496 luHieGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuG 516
Db : : : : :
3875 ACTACGAGGAGG-----CGGCCGCTTATGATNAACTCG 3910
QY 516 luThrSerLysSerValGluAlaAlaValleArgIleThrCysSerArgThrSerThrg 536
Db : : : : :
3911 AAAAGACCAAG-----AACAGGCTTCAGCAGG 3937
QY 536 luValGlyAspLeuLysLeu---LeuAspHis-----A 546
Db : : : : :
3938 AGCTGACGACCTGGTTGTGATTTGGACAACAGCGGCACTCGTGTCCAACCTGGAAA 3997
QY 546 snTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuG 566
Db : : : : :
3998 AGAAGCAGAGGAAATTTGATCAGTTGTTAGCCGAGGAGAAAACATCTCTCCAAATACG 4057
QY 566 luGlnMethHisAspValAlaLysLysLeuAlaGluAspGlu 580
Db : : : : :
4058 CGGATCAGAGGAGCAGCTGAGGCAGAGCCAGGAGGAGGAA 4101

RESULT 13
US-09-949-016-1240
; Sequence 1240, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1240
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1240

Alignment Scores:
Pred. No.: 2,66e-05 Length: 6861
Score: 144.00 Matches: 133
Percent Similarity: 35.38% Conservative: 92
Best Local Similarity: 20.91% Mismatches: 216
Query Match: 4.48% Indels: 197
DB: 4 Gaps: 28

SQ03_D1219_COPY_650_1270 (1-621) x US-09-949-016-1240 (1-6861)
QY 22 GluLysIleLysValGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db : : : : :
2987 CAGAGCTGCACTTGAAGGTCACGGCTGAGGCCAAGATCAAGAACTG----- 3037
QY 42 CysThrValGluGluLysAlaLeuIleLysAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db : : : : :
3038 -----GAGGATCAGATCCTGGTGTGATGATCAGAACAAATAAATATCAAAA----- 3085
QY 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db : : : : :
3085 ----- 3085
QY 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
Db : : : : :
3086 -----GAA 3088
QY 102 GlnSerIleMetAspGlyValleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 12:27:00 ; Search time 983.333 Seconds
(without alignments)
3964.922 Million cell updates/sec

Title: SEQ3_D1219_COPY_650_1270
Perfect score: 3217
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/ROBINSON929/runat_11072005_102153_29226/app.query.fasta_1.2325
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=DOCALLIGN-p2o -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000
-USER=ROBINSON929 @CEN 1.1 1320 @runat_11072005_102153_29226 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US07_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	590.5	18.4	6409	19	US-10-437-963-76223	Sequence 76223, A
2	495.5	15.4	8425	19	US-10-437-963-69891	Sequence 69891, A
3	456.5	14.2	5865	19	US-10-437-963-76215	Sequence 76215, A
4	335.5	10.4	1197	20	US-10-425-115-159032	Sequence 159032, A
5	192.5	6.0	487	18	US-10-424-593-12966	Sequence 12966, A
6	164	5.1	7764	10	US-09-814-353-21548	Sequence 21548, A
7	164	5.1	7764	17	US-10-341-434-60	Sequence 60, Appl
8	164	5.1	7764	21	US-10-651-237-27	Sequence 27, Appl
9	164	5.1	7764	21	US-10-782-413-27	Sequence 27, Appl
10	161	5.0	661	20	US-10-425-115-145108	Sequence 145108, A
11	160	5.0	3045	17	US-10-369-493-26897	Sequence 26897, A
12	158	4.9	5883	21	US-10-956-157-859	Sequence 859, Appl
13	158	4.9	7274	21	US-10-956-157-859	Sequence 48, Appl
14	158	4.9	7396	17	US-10-028-248A-35	Sequence 35, Appl
15	158	4.9	7396	17	US-10-107-782-35	Sequence 35, Appl
16	157	4.9	6354	15	US-10-084-817-158	Sequence 158, Appl
17	156.5	4.9	4551	18	US-10-675-072A-42	Sequence 42, Appl
18	154	4.8	5119	21	US-10-956-157-4967	Sequence 4967, A
19	154	4.8	5857	10	US-09-873-367C-305	Sequence 305, Appl
20	154	4.8	5857	21	US-10-843-641A-305	Sequence 305, Appl
21	154	4.8	5857	21	US-10-831-704-79	Sequence 79, Appl
22	154	4.8	5857	21	US-10-956-157-975	Sequence 975, Appl
23	154	4.8	5857	21	US-10-956-157-975	Sequence 975, Appl
24	154	4.8	6364	13	US-10-071-766-140	Sequence 140, Appl
25	153.5	4.8	6965	20	US-10-335-053-229	Sequence 229, Appl
26	153.5	4.8	7694	16	US-10-096-534-34	Sequence 34, Appl
27	153.5	4.8	7695	11	US-09-968-007A-462	Sequence 462, Appl
28	153.5	4.8	7695	21	US-10-843-641A-462	Sequence 462, Appl
29	153.5	4.8	8063	10	US-09-814-353-21776	Sequence 21776, A
30	151	4.7	2434	20	US-10-425-115-109076	Sequence 109076, A
31	151	4.7	7659	19	US-10-437-963-27251	Sequence 27251, A
32	149.5	4.6	5847	16	US-10-032-585-6611	Sequence 6611, A
33	149	4.6	5641	16	US-10-032-585-6646	Sequence 6646, A
34	145.5	4.5	5676	19	US-10-437-963-58717	Sequence 58717, A
35	144.5	4.5	4944	16	US-10-032-585-6572	Sequence 6572, A
36	144.5	4.5	10096	9	US-09-960-253-163	Sequence 163, Appl
37	144.5	4.5	10096	17	US-10-172-118-1127	Sequence 1127, A
38	144.5	4.5	10096	18	US-10-342-887-1127	Sequence 1127, A
39	144.5	4.5	10096	21	US-10-848-755A-138	Sequence 138, Appl
40	144.5	4.5	10326	20	US-10-751-736-36	Sequence 36, Appl
41	144	4.5	2648	19	US-10-437-963-85842	Sequence 85842, A
42	144	4.5	3388	9	US-09-954-456-1602	Sequence 1602, A
43	144	4.5	3388	9	US-09-967-768A-245	Sequence 245, A
44	144	4.5	3388	9	US-09-954-531-988	Sequence 988, A
45	144	4.5	3388	9	US-09-954-531-1382	Sequence 1382, A

ALIGNMENTS

RESULT 1

US-10-437-963-76223
; Sequence 76223, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barzak, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

With Other Molecules Associated

Alignment Scores:				
Pred. No.:	1.6e-50	Length:	6409	
Score:	590.50	Matches:	175	
Percent Similarity:	45.21%	Conservative:	127	
Best Local Similarity:	26.20%	Mismatches:	239	
Query Match:	18.36%	Indels:	127	
DB:	19	Gaps:	20	
SEQ3_D1219_COPY_650_1270 (1-621) x US-10-437-963-76223 (1-6409)				
Qy	3	AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu	22	
Db	2263	GATGCCATCATCATATATATAGTAGATGGAAATCCAAACCAATGACTTTGAGAGTTCACG	2322	
Qy	23	LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys	42	
Db	2323	AGGATCAGTATAGAGTCTCAATCTGAATGTGTGCTATTTTCGTTGTATTCGTTGT	2382	
Qy	43	ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal	62	
Db	2383	ACGGTGAGGAAAGACTCTTATACTAGCGAAGCATGATCATATATCTTGACAGCAAGCTC	2442	
Qy	63	GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe	82	
Db	2443	CAGAATGTAATGCCATCTATCTGAGCCATTCCTGTAGTTGGGTGCATCGTTCTCTTC	2502	
Qy	83	AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly	97	
Db	2503	AATAGACTTGAGGATTCCAAAGCATGACTATCCAGTAAGATTCGAGGATGATGGT	2562	
Qy	98	ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer	117	
Db	2563	CTC-----TTCATCAATAATGATATTTTGGAGTTTGAGCGCAACTATATCC	2607	
Qy	118	SerLysGlyGlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAla	137	
Db	2608	ACCAAGCTTGAAGCCAGCACATAAATGGAAATGCAGTCATATCTCAGGCTCAGCAGAT	2667	
Qy	138	GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu--AspHisIleLysLeuSer	156	
Db	2668	GGATCGTTTATTCTACAGATATTCGTGTTATTAACGAGAGGGAGGAATATCTGCAGTT	2727	
Qy	157	AspGluGluSerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnPro	174	
Db	2728	GATCGTGATTTGCCAAGATTCAGACCTCTCTGGTCAAAATTTACTAGTGGGAGTCTCCT	2787	
Qy	175	MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu	194	
Db	2788	CATTGGCAATATATATCTGAGCCAGTGCAAAGGAACCGTAGGAAGATACAAAATATGAA	2847	
Qy	195	GlySerGluAlaSerProLysThrGlyAspGlyClyAsnAlaLysLysArgLysLysAla	214	
Db	2848	GACCAATAGAAATACCTCGCAGAGAAGAACTGATGAAGCAATATGAACCTGAGAAAAATC	2907	
Qy	215	SerAspAspValThrAspProArgValThrAspProProValAspAspGluArgLys	234	
Db	2908	GGAGAA-----ATCATGGAT-----	2922	
Qy	235	AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSer	254	
Db	2923	-----TCATCTCCGAAGATCTCTACCTGCTCAAGAC	2952	

QY 579 AspGluAlaCysTrpIleAsnArgIleLys-----SerTrpAlaAlaLys 593
Db 3892 CACATGAGATGCCAGCAACAAGTTGAAGATGACAGCAATCAACTTCATGGAACAAGGAG 3951
QY 594 LeuLysValCys---ValProIleGlnSerGlyAsnAsnLysHisPheSerGlySerSer 612
Db 3952 TTACAATTGAAGGAATTTTCTTCCAGCAAGCAAAATCTGGCCACTTAGATCGTAGCTTT 4011
QY 613 AsnIleSerGlnAsnAlaProAsp 620
Db 4012 GATCAACAATAATCTCTTACCAGAT 4035

RESULT 2

US-10-437-963-69891/c
; Sequence 69891, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69891
; LENGTH: 8425
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRF4530_70516C.1
US-10-437-963-69891

Alignment Scores:

Pred. No.: 2, 41e-40 Length: 8425
Score: 495.50 Matches: 176
Percent Similarity: 40.69% Conservative: 108
Best Local Similarity: 25.21% Mismatches: 255
Query Match: 15.40% Indels: 159
DB: 19 Gaps: 20

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-437-963-69891 (1-8425)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 5785 GATGCCATCATCATATATGTTCTGACAAACCCACTAAACGACTTGAAGGCTCTTCAA 5726
QY 23 LysIleIleLysLeuGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 5725 AAGATCAAGATAGAGTCACAAATTTGACGCGTGTGAGCAATTTTCGCTGTACACTCTCTTT 5666
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 5665 ACAGTTGAGAGAGAGAGCTTGTCTTTCGAGGCGGCGGCAATTGTTATTGATAACAACATA 5606
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 5605 CAAGATCTAAGGACTAGTTTAAACATTCCTTACTCAGATGGGCTGCTGCAATTCCTCTTC 5546
QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
Db 5545 AGTAGATTGATGGTCCAGCAAGATGACCATGCAAGTAAAGTTCTGAAATGGGAGAGA 5486
QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGlyGlyGlu 122
Db 5485 CACTTTATTGATGAAGTAATCGTTGAGTTCTTTAAACAAAGTTATCCACACTGTTGAAGAC 5426

QY 123 GluAsnGluValLysLeuCysLeuLeuLeuAlaLysHisAlaGlnGlyThrTyrSer 142
Db 5425 AGCACTGAAGTACACAGCAAGTCCATAAGCAAGCTAATATGATGGGGAACATATATCA 5366
QY 143 SerAspSerThrLeuPheGlyGluAspHisIleLeuLysLeuSerAspGluGluSerProAsn 162
Db 5365 AGAAATATTACTCTGATGGTGAAGAGGGAATCTCTGCTGTGAGGATAATCCAGCT 5306
QY 163 IlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAspThr 182
Db 5305 GAGTCTCTGTTAAATTTATTGATGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 5246
QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluLysSerProLysThr 202
Db 5245 CTCGAATCAAGATTACAAGTCCCAGACCATGAT-----GAGTCAATGCTCCGGCT 5192
QY 203 GlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspVal-----Thr 219
Db 5191 GAAGAATCAATGAAGCCAGAAAGACATAGAAAGGTTGGTGAGATTATGGGTTTCATCT 5132
QY 220 AspProArgValThrAspProValAspAsp-----AspGluArgLysAlaSer 236
Db 5131 TCAAAAAGTTGATCTGACAAAAGTAACGATGATGCTGCTGATATTTGTACTACATCT 5072
QY 237 GlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 5071 GGGCCCT-----GCTCTTCAA---CCAGTTGATGTCACGCAAGCAAAATCAAGT 5037
QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
Db 5026 CAATCTGAAGGTCGGAAGCCTGATG-----5000
QY 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
Db 5000 -----5000
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4999 -----AGTACACCCCAAGAAATCTTCATGCCCAAAATGAAGCAAGTGTGCAAAATTAAT 4946
QY 317 GlnValLeuHisLeuSer-----322
Db 4945 AAGGTGTTACACTACCAATACCAATACAGTTTAATGGAGCGAAATAGCAGGACAGTT 4886
QY 323 -----AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
Db 4885 CTCATTGGAAGTACACAATGTTACGTTGCTGCTGAGCAGTCTCTTGAATATCTTTGAAT 4826
QY 339 AsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSer 358
Db 4825 AATCATGTAGTTGTCAGGAGCCAAAGTACATATTTTCATGCGTTGAACATAGCTTTGTGT 4766
QY 359 TrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValAlaGala 378
Db 4765 TGCGAGTGGCTCTTATTCATACTTAAGTGGATCATTAAGAGTCACCTTGCCCTTGCT 4706
QY 379 AsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398
Db 4705 GAAAAAGGTTAAAAATACGAGTCAATGAGGAGCTGCAAGGTTGGTCTATGACAGCCTA 4646
QY 399 SerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGly 418
Db 4645 -----AAGAGAAAAATTTCCAAAAGAAAGCAGGTCGAACCTGGCAGCAATTTGTCATCA 4595
QY 419 ThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAla 438
Db 4594 ACA-----TCAGTAGAAAAGACCAACCTTCACAGCAGGAGACCTCTAAT--- 4550
QY 439 ThrValArgAspGluLysIleAsnThrLys-----448
Db 4549 ATCTTGAGAAATGACCCACATTTTCCCTAAGCAGAGAATGGATCTTCATGATAATTTTCATG 4490
QY 449 -----SerMetArgAsnSerSerGluAspGluGlu 458

Db	4489	AATGGTCAC	TTCAAGAAGGCTCA	TTGTTGCTGCACAAATGGTTTCAGAAAGCAGGAA	44300	
		: : : : : :		
Qy	4459	CysMet	-----ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp	474		
		...				
Db	4429	CTGATAGCTG	TTCCAGGAACCTCACATGGAAATGC	--CATTTTCAACTGATGAAC	TTCTCT	4373
Qy	475	-----l	leGluLysThrIleSerGlyLeuLysLysTyrLysLysGlnValGlnLys	492		
Db	4372	GACATTTGGG	AAAGAAGATCAACTTAATGACAAATGTTTCTCTGTAAGAGGAATACCGT	4313		
Qy	493	LeuValGlnGluHisGluGluLysMetGluLeuLeuAsnMetTyrAlaAspLysLys	512			
				
Db	4312	ATCTTTGATAAA	CAACAGCTCTCAGATTCAGAGTTAGAGAAATACACACAAAATAAAACT	4253		
Qy	513	GlnLysLeuGluThrSerLysSerValGluAlaValIleArgIleThrCysSerArg	532			
Db	4252	GCCAGACTG	AAACACA-----GTGTGCAAT	--	4229	
Qy	533	ThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAsp	552			
Db	4228	-----TTAGT	TCTCGAAATATATTTGTAGAGTCATGCTGAT	4193		
Qy	553	GluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisAspValAla	572			
Db	4192	-----GTAGAG	ACCCGAAATGACACAAATAAAACAACTGTTCACGTGGTTCACTATGCTG	4139		
Qy	573	LysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSer	589			
Db	4138	ATGTATGCAT	TTCTGGAGCATCAGACTCCAGCACAGCAAACTTTAGTCTACTGCAATCT	4079		
Qy	590	-----TrpAla	-----AlaLysLeuLysValCysValProIleGlnSerGly	603		
Db	4078	AATACATGGG	CTGAAGCGCACAGCTGAAGGAAAACTTTGTTTGGGAACAAATACAGC	4019		
Qy	603	-----	-----	-----	603	
Db	4018	CAATTAGAT	CACCTTTGATCAGACAGATGCTCTACAGAGTTCAAACTTTGTTATGCA	3959		
Qy	604	-----	-----AsnAsnLysHisPheSerGlySerSer	612		
Db	3958	GAATTCAT	CCATTTGAAGGAACAAAGTAGCAATCCCATGTTGCTGCTGAAGCGCT	3905		

```

RESULT 3
US-10-437-963-76215
; Sequence 76215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76215
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1
US-10-437-963-76215

Alignment Scores:
Pred. NO.: 1.77e-36
Score: 456.50
Matches: 160
Length: 5865

```

```

Db 3007 CACACCAACTCAGCCAGAGCTGTCAAAGTTGTATGAATTTGGTAATTCGCGGAGACT 3066
Qy 325 CysThrSerMetValGluAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGlu 344
Db 3067 GTCAAAATGTTTATGTAGGAATTAATGATTAATTTGAAGATCATCAA----- 3117
Qy 345 GluProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTyrPheAlaAlaLeuLeu 364
Db 3118 |||||-----TGTTCGCGTGTCTCTCTC 3138
Qy 365 ValIysGlnIleLeuSerHisIleValSerLeuValArgAlaAsnSerGluLeuAlaPhe 384
Db 3139 TTAAGCATATGATCATCGAGAGTCACTTCCCTTCTGTGAGAACTTGAATAT 3198
Qy 385 LysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeu 404
Db 3199 GAGTGCATGAAGTCTCGTGAATATGTTTATGAGAAGCTAAGGATCTCAAGAAAAG 3258
Qy 405 PheLeuGluHisThr-----GlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 3259 TTTTCGCGTAGACCAAGTGAACCAAGCAAGCAGAGCCCAATCTACTCCAGTAAATAACACA 3318
Qy 420 AsnSerLysGlnSerValSerThrLysLeuValAsnGluSer-----LeuSerGly 437
Db 3319 TCATCTTACAAACACAGACTCACCGAATTTGAGAGTACGAGTCAATTTGTCTATCAG 3378
Qy 438 AlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAsp--- 456
Db 3379 GTAACGACAATTTGATGTGTATTTGGAAAAATGTTTCACATCAAGAGGCTCCACATGATC 3438
Qy 457 -----GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 3439 TTGACCGAGAGATGATATTAGAACAGAG----- 3468
Qy 475 IleGluLysThrIleSerGlyIleLysLysLysTyrLysGlnVal-----GlnLys 492
Db 3469 -----GAATGATATCTGTTCTAGAAATCATAGAGAAGACATGTTTAAAGGATGAA 3522
Qy 493 LeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLys 512
Db 3523 CTTCCTTGAAAGATCACAGAGAAAGAAATTAATTAATTAACATGTTTCTCCTTAAGA 3582
Qy 513 GlnLys-----LeuGluThrSerLysSer 520
Db 3583 GAAAGAAGATATTCAGGATAAACCAAGCAAGCAAGCAGACCACTGTTAGACATGCAACAGCA 3642
Qy 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540
Db 3643 AAGAGTGGCAAGCTGCAAGAAACATGCAAT----- 3675
Qy 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
Db 3676 TTGTCGTGGACATCTTCGTAAAGTGCATTTGAT-----TCAGAGGACAGGGATGCC 3729
Qy 561 CysLeuLysSerLeuGluGlnMetHisAspValAlaLysLysLysLeuAlaGluAspGlu 580
Db 3730 ACGGTGAAGTAATAATTAATGATGTTTACTTCTGCTGATGCAATTTTGAATCATCATG 3789
Qy 581 AlaCysTrpIleAsnArgIleLys-----SerTrpAlaAlaLysLeuLys 595
Db 3790 AGATGCCAGCATACCAAGTTGAAGATGCAAGCATCAACTTCTATGGAACAGGAGGTATCAA 3849
Qy 596 ValCys---ValProIleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIle 614
Db 3850 TTGAGGAATTTTCTTCTAGCAAGCAAAATCTGGCCACTTAGATCGTACCTTTGATCAA 3909
Qy 615 SerGlnAsnAlaProAsp 620
Db 3910 CAAATTCCTTTACCAGAT 3927

```

RESULT 4

US-10-425-115-159032

; Sequence 159032, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 159032

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(1197)

; OTHER INFORMATION: unsure at all n locations

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_76611C.1

; US-10-425-115-159032

Alignment Scores:

Pred. No.:	8,54e-25	Length:	1197
Score:	335.50	Matches:	102
Percent Similarity:	48.63%	Conservative:	58
Best Local Similarity:	31.00%	Mismatches:	120
Query Match:	10.43%	Indels:	49
DB:	20	Gaps:	12

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-425-115-159032 (1-1197)

```

Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 142 GATGCCATCATATATTTGTAGTGAATCCCATGAATGACITTGAGGCCCTTCAG 201
Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 202 AGGGTCAGCATGAGTCAGACTCTGAACCTGTGCTATTTTCTGTTGTACTCGTCTTT 261
Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 262 ACAGTGGAGAAAGCTCTTATTTCTGCAAGACATGACCATATTTTGGATAGTAATATT 321
Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuGlu 82
Db 322 GTCAATATACCCCTAGCCTGACCATTTGCTCTCAGTTGGGGCGCATCTTCTCTTC 381
Qy 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
Db 382 AATAGACTTGAGGAGTTGCAGCAACATAGCTAC-----TCAATGCTCTGCT 429
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 430 GATGAGCTTTTCATGATAATGTAGATTGAGAGTTTTCAGCAAAATTTACTCAGCAAGTT 489
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 490 GAACTCAGAACTGAATCAGGCAATACAGCCATATCAGACTTATCTGTGTGATCTTTC 549
Qy 141 TyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGluGlu 159
Db 550 TATTTCTAGGCTATTGTTGTAGCGGTGAGAGGAGGAAATTCATCAGTTGATGGTGAC 609
Qy 160 SerProAsnIle-----PheTrpSerLysLeuLeuGlyLysAsnProMetTrpLys 177
Db 610 CTGCCAAAGTTCTGGGCTACTGGGTAAAGTTTACTTAATGGGAGGTCTCTCAGTGCAG 669
Qy 178 TyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGlu 197
Db 670 TATATATCTGAGCCAGTACAAAGAGCCGTAGAAAGATA-----AATAACATGGAA 720

```

```
QY 198 AlaSerProLysThrGlyAspGlyGlyAenAlaLysLysArgLysLysAlaSerAsp 217
Db 721 CAACAATTTGAAAACACAGAC-----AAACAGTTGAAAATAAATACTACTGAAGAA 768
QY 218 ValThrAspProArgValThrAspProProValAspAspGluArgLysAlaSerGly 237
Db 769 ACTGATGAAGCCAGAGTA-----AAGCGTAGAAGAATTGGT 804
QY 238 Lys---AspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 805 GAATCATGATCATCTGCAATAGACTCACCT----- 837
QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
Db 838 ---GGCAAAACAAAGACACCATATATGCGCAGGAACAATACC-----CCACCT 882
QY 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
Db 883 AGTTCTCAT-----CAATATCATGTTGAAGATACATGCGCAGGAACATA 924
QY 297 Pro-----AspGluSerGlnArgAlaGluHisThrValLeuLysProLysMet 312
Db 925 GAGCGAAGCAATCTTCATCCACACAAAAGGCTACACGTTCACTCACTCAAGCGAGATT 984
QY 313 AlaLysLeuCysGlnValLeuHisLeu 321
Db 985 TCGAAGTTATATAAGTTGCTGCAACTG 1011
RESULT 5
US-10-424-599-12966
; Sequence 12966, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 12966
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1
US-10-424-599-12966
Alignment Scores:
Pred. No.: 2,43e-10 Length: 487
Score: 192.50 Matches: 57
Percent Similarity: 45.64% Conservative: 32
Best Local Similarity: 29.23% Mismatches: 71
Query Match: 5.98% Indels: 35
DB: 18 Gaps: 4
SEQ3_D1219_COPY_650_1270 (1-621) x US-10-424-599-12966 (1-487)
QY 337 IleGluAsnHisArgIleTyrGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 1 ATGAACAATCTTTATGTCATAGGGAACCGTTTCAATATATACAGGCTTTTCAATATCT 60
QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 61 CTGTGTGGACGTGCAGCTTCTTGTGTAACACAAACTTTGACCCCATCTTCTCTCTATA 120
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSer 396
Db 121 CAA-----GATTTGAACCTTTGAGTGTAAGAAGAAGAGGTGGACTACATTTGTTCT 171
```

```
QY 397 IleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCys 416
Db 172 ATGTTGTGTTGTTTGAAGAAAATAATCTTATATCGCACAGGAATAATATCATGATACCGGC 231
QY 417 PheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSer 436
Db 232 TTT----- 234
QY 437 GlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAsp 456
Db 235 -----ACAAAAGCATCTGGACCATCAACACGGGCA 264
QY 457 GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGlu 476
Db 265 TATTCCTGTACAGGAGTTGCACGA-----GAGGTTGAATTGTTCAAAAAGATATGTCC 318
QY 477 LysThrIleSerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGlu 496
Db 319 AAAAGTATTAAAGAAATTCAGAAAGAGTGTGAAAAGAGCTGAAGAAGCTACATATTTTG 378
QY 497 HisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGlu 516
Db 379 CAAGAAGAAGAGAACAAAGGTTGAGACGAGCTATTGAGGAAGAAAAGCGTAATTTTGAT 438
QY 517 ThrSerLysSerValGluAlaValIleArgIleThrCysSer 531
Db 439 GAAAGGTACAAACATAGAGTCTGCTGATATTCGA---TCTGTCTCT 480
RESULT 6
US-09-814-353-21548
; Sequence 21548, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21548
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 7755, 7756, 7757, 7758, 7759, 7760,
; LOCATION: 7761, 7762, 7763, 7764
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21548
Alignment Scores:
Pred. No.: 1.64e-05 Length: 7764
Score: 164.00 Matches: 162
Percent Similarity: 33.87% Conservative: 112
Best Local Similarity: 20.02% Mismatches: 252
```

Query Match: DB:	5.10% 10	Indels: Gaps:	283 37
<hr/>			
SEQ3_D1219_COPY_650_1270 (1-621) x US-09-814-353-21548 (1-7764)			
<hr/>			
Qy	2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHieVal 21		
Db	3342 GCGACACAGCTGGTCATCTTTTGACTCTGCAGTCCCAAGAACCCCATGAATGTTCAGGCCAA 3401		
<hr/>			
Qy	22 GlutylLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTySerVal 41		
Db	3402 GCCGAGCGCATAGAATTGGTCAGAAGACGAGTAATAATTACCCTTAGTTACAAG 3461		
<hr/>			
Qy	42 CysThrValGluGluLysAla-----Leu 49		
Db	3462 GGGAAGCTGGAGGAGGAGCATATAGAACGGGCCAAAAAGAAAGATGGTATTAGATCATCTG 3521		
<hr/>			
Qy	50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn----- 64		
Db	3522 GTGATTTCAGCGCATGCACACCATCGGGCGGACGATCCTGGAAAACAACACTCAGAGAGTCC 3581		
<hr/>			
Qy	65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80		
Db	3582 AACTCAAATCCTTTTAATAAGAGAGCTGCACGCTATTTTGAATTTGGCAGCAGGAT 3641		
<hr/>			
Qy	81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100		
Db	3642 CTCTTCAAGAACTGGAA-----GGGAGGAATCA 3671		
<hr/>			
Qy	101 GluGlnSerileMetAspGlyValIleHisgluPheSerSerileLeuSerSerLysGly 120		
Db	3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTTCGGTGGCGTGAA 3716		
<hr/>			
Qy	121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140		
Db	3717 ACAGAGAGAAATGAAGTGTCAACAAGTGCACAGATGAATCTTA---TCACAGTTTAAG 3773		
<hr/>			
Qy	141 TyrSerSerAspSerThrLeuPheGlyClyAspHisileLysLeuSerAspGluGluSer 160		
Db	3774 GTTGCCAACCTTGCACAATGGAAGATGAAGAAGAGCTA-----GAAGAGCGT 3821		
<hr/>			
Qy	161 ProAsnillePheTrpSerLysLeuLeuGlyLysAsnPrometTrpLysTyrProSer 180		
Db	3822 CCTCACAGNACTGGGATGAGATCAT------- 3848		
<hr/>			
Qy	181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu----- 194		
Db	3849 -----CCAGAGGAACAAGAAAGAAAAAGTAGAGGAGGAAGCGGCAGAGGAGCTAGAA 3902		
<hr/>			
Qy	195 -----GlySerGluAlaSerProLysThrGly 203		
Db	3903 GAAATTTATATGCTGCTCGAATTGGAGTTCACATAAAGGCTCAGACAAATGACAGT 3962		
<hr/>			
Qy	204 AspGlyGlyAsnAlaLysArgLysAlaSerAspAspValThrAspProArgVal 223		
Db	3963 GACTCTGACACTGAGCTAAGAGCGCCAGAGATCCTCTGCTTCTGAGAGTGAAACG 4022		
<hr/>			
Qy	224 ThrAspProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243		
Db	4023 GAAGAC-----TCTGATGATGACAAGAAG----- 4046		
<hr/>			
Qy	244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263		
Db	4047 -----CCAAAGCCGACAGGGCGTCCGAGGAGTGTGCGGAAGGACCTCTGTGGAGGGA 4097		
<hr/>			
Qy	264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeutySerMetGlySer 278		
Db	4098 TTTTACTGATGCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT----- 4145		
<hr/>			
Qy	279 HistleSerGlylleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296		
Db	4146 -----GGTCTCCCCTTTGAACCGCTGAGTGCAATGACGCTGAT--GCTGAGCTG 4193		

Qy	297	ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetalAlaLeuCys	316
Db	4194	GTACATAAGTCGGTGGCAGATCTG-----AAGCGCCTGGGT	4229
Qy	317	GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal	336
Db	4230	GAACCTGATCCAC-----AACACGCTGTGTCTACCAATGCAGGAATACGAAGAGCAGCTG	4283
Qy	337	IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla	356
Db	4284	AAAGAAATATCCACGAGGAGGAAAAGGACCAAGGAAAAGG---AGAGTCCCAACATCAAG	4340
Qy	357	LeuSerTrpIleAlaLeuValLeuValGlnIleLeuSerHisLysGluSerLeuVal	376
Db	4341	ATATCCGGAGTT---CAGGTTAATGTGAAATCCATTATCCAAATGATGAAGAGGAGTTT---	4394
Qy	377	ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle	394
Db	4395	-----GAGATGCTGCATAATCATCCCTGTGGACCCCTGAAGAAAAA-----	4442
Qy	395	TyrSerIleLeuSerCysMetLysSerPheLeuGluHisThrGlnGlyLeuGlnPhe	414
Db	4443	TACTGCTTAACCTGCTCTGTCAAAAGCTGCACATTTTGTATAGTGGGGGTGGAGAT	4502
Qy	415	AspCys-----PheGly-ThrAsnSer-----Lys	422
Db	4503	GATTCTCGCTGTGCTGGGATTTATGAACATGGCTATGCAAACTGGGAGTTAATTAAA	4562
Qy	422	sGlnSerValValSerThrLysLeuValAsn-----	432
Db	4563	ACAG-----ACCAGAGCTTAAATTAAC TGACAAAATTCGCGGTGGAGACA	4610
Qy	433	----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet	450
Db	4611	GATAAAGAGCTCAGGGAAGAGCTACAGCCGAGCGGATTTACTTGTGTGAAGTGTCTC	4670
Qy	451	ArgAsnSerSerGlu-----AspGluCysMetThrGluLys	463
Db	4671	AGAAAGGCTCTGGAGAAAGAGGGGGCTGTGCAGGTGGGGAAGAGGCCAAATTAAGAAG	4730
Qy	464	Arg-----	464
Db	4731	CGGAAGCCTCGGGTAAAGAAAGGAAAAACAAGTCCCAGGCTGAAGAGGAGCATGGAATT	4790
Qy	465	-----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
Db	4791	GAGCTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGGAGNAAGTGAAGATGAT	4850
Qy	481	GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
Db	4851	GGCTTGGAAAAAGTCCCAATGAAAAA-----ACAGAG---AAGAAAGAGAACAAAGGAGAAC	4907
Qy	501	LysMetGluLeuLeuAsnMetTyrAlaAspLys-----	511
Db	4908	AAGGAGAAACAAATAGATTCTTAGGAAAGACAAAGAAAGGGGCAAGGAAAGAAAGTCA	4967
Qy	512	LysGlnLysLeuGluThrSerLysSerValGluAla-----	523
Db	4968	AAAGATAGAAAGAGAAAGCCCTAAAGTGTGTATGCCCAATCTTCGAGTAATCAAGCGGA	5027
Qy	524	-----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----	539
Db	5028	TCTCAGGGTCTGTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGAGGATGAG	5087
Qy	539	-----	539
Db	5088	GATGATGATCTGGACCAGGAGACATTCAGCATATGTAAAGGAGAGGATGAGGCCCGTGAAA	5147
Qy	540	-----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp	552
Db	5148	AAGGCATGAAACAGCTCGCAAAACCTCGAACAGGGGCTCACTGTCGAAGAACAGCTGGAA	5207
Qy	553	-----GluIleLysSerGluLysAsnGluCysLeuLysSer---	564

```
Db      5208 CACACCGGAAGTCTGCTGAAATATCGGAGACCGGATAGCCGAGTGCCTTAAGCGCTAC 5267
Qy      564 -----|||||-----
Db      5268 TCAGATCAGGAGCACATCAAACTCTGGAGAGAACTATGATTTTGTTCCTCAAGTTT 5327
Qy      565 -----LeuGluGlnMetHisAspValAlaLysLysLysLysLeuAla 577
Db      5328 ACAGAAATTTGATGCTCGAAAATCGCATAGTTATACAGATGGCTCATAGAAAGGTTCT 5387
Qy      578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLysValCys 597
Db      5388 CAAGAAGAAGAG-----GAGCAAAAGAAGAAGACGAC 5420
Qy      598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db      5421 GTG-----ACTGGGGGTAAAGAAACCATTTTCGTCCAGAGCGCTCAGGCTCCAGCGCG 5471
Qy      614 -----IleSerGlnAsn 617
Db      5472 GACTCTCTGATATCTCAGTCC 5492

RESULT 7
US-10-341-434-60
; Sequence 60, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (708)..(5924)
; OTHER INFORMATION:
US-10-341-434-60

Alignment Scores:
Pred. No.: 1,64e-05 Length: 7764
Score: 164.00 Matches: 162
Percent Similarity: 33.87% Conservative: 112
Best Local Similarity: 20.02% Mismatches: 252
Query Match: 5.10% Indels: 283
DB: 17 Gaps: 37

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-341-434-60 (1-7764)
Qy      2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
Db      3342 CGGACACAGTGTCTATCTTTGACTCTGACTGGAAACCCCAAGATCACTTGCAGGCACAA 3401
Qy      22 GluLysIleLysIleGluSerCysSerGluArgThrLysAlaPheArgLeuTyrSerVal 41
Db      3402 GCCCGAGCGCATAGAAATTTGGTCAGAAGAGCAGGTAAATATTTACCGCTTAGTTACAAAG 3461
Qy      42 CysThrValGluGluLysAla-----Leu 49
Db      3462 GGGACTGTGGAGGAGGAGATCATAGAACGGCCAAAGAGAGATGTTATTAGATCATCTG 3521
Qy      50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn-----64
```


QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
|||
Db 4443 TACTGCTTAACCTGCTGTCACAAAGCTGCACATTTTGTATAGAGTGGGGTGGAGAT 4502
|||
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
|||
Db 4503 GATTCGCTGTTGCTGGGGATTATGAACATGGCTATGGAACCTGGGAGTTAATAA 4562
|||
QY 422 sGlnSerValSerThrLysLeuValAsn----- 432
|||
Db 4563 ACAG-----ACCCAGAGCTTAATTAACACTGACAAAATTCGCCGGTGGAGACA 4610
|||
QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
|||
Db 4611 GATAAAAGCCTCAGGGAGACGCTACAGACCGGAGCGGATTACTTGTGTAAGCTGCTC 4670
|||
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
|||
Db 4671 AGAAAGGCTCTGGAGAGAAAGAGGGGCTGTGACAGGTGGGGAGGAGGCCAAATTAAGAAG 4730
|||
QY 464 Arg----- 464
|||
Db 4731 CGGAAGCTCGGTAAAGAGAAACAAAGTCCCGAGCTGAAAGAGGAGCATGAAT 4790
|||
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
|||
Db 4791 GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAGAGGAGAGTGAAGAATGAT 4850
|||
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
|||
Db 4851 GGCTTGGAAAAAGTCCAAATGAAAAAACAACAAG---AAGAAAGAGAACACAGGAGAAC 4907
|||
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
|||
Db 4908 AAGAGAACAAATAGTTCTAGAAAGACAAAGAGAGGAGGACAAAGAAAGAAAGATCA 4967
|||
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
|||
Db 4968 AAGATAAAGAAAGAGAGCCATAAAGTGGTGTGTCGCCAAATCTTCGAGTAATCAAGCGA 5027
|||
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
|||
Db 5028 TCTCAGGGTCTCTCCATATTACAGCAGGAAGTGAACCTGTCTCCCATTTGGAGAGGATGAG 5087
|||
QY 539 ----- 539
|||
Db 5088 GATGATGATCTGACACGAGAGACATTCAGCATATGTAAAGAGAGATGAGGCCCGGTGAAA 5147
|||
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
|||
Db 5148 AAGGCCTGAAACAGCTCGACAAACCTGACAGGGGCTCACTGTCAGAACAGCTGGAA 5207
|||
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
|||
Db 5208 CACACCGGAACCTGCTGCTGAAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5267
|||
QY 564 ----- 564
|||
Db 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGATTTTGTTCACAGTTT 5327
|||
QY 565 -----LeuGluGlnMetHisAspValAlaLysLysLysLeuAla 577
|||
Db 5328 ACAGATTTCATCTCGAAACCTGCATATAGTTATACAGATGGCTCATAGAAAGAGTCT 5387
|||
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
|||
Db 5388 CAAGAAGAAGAG-----GAGCAAAAGAAAGAAAGACGAC 5420
|||
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
|||
Db 5421 GTG-----ACTGGGGGTAAAGAAACCAATTTCTGCCAGAGGCTCAGGCTCCAGCCGG 5471
|||

QY 614 -----IleSerGlnAsn 617
|||
Db 5472 GACTCTCTGATATCTCAGTCC 5492
|||

RESULT 8

US-10-651-237-27
; Sequence 27, Application US/10651237
; Publication No. US20050048494A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: ADS-5003 US NP
; CURRENT APPLICATION NUMBER: US/10/651,237
; CURRENT FILING DATE: 2003-08-27
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
US-10-651-237-27

Alignment Scores:
Pred. No.: 1,64e-05 Length: 7764
Score: 164.00 Matches: 162
Percent Similarity: 33.87% Conservative: 112
Best Local Similarity: 20.02% Mismatches: 252
Query Match: 5.10% Indels: 283
DB: 21 Gaps: 37

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-651-237-27 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
|||
Db 3342 GCGACACAGTCGTCTATCTTGTACTGTGACACCCCAAGATGACTTGCAGGCACAA 3401
|||
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
|||
Db 3402 GCCCGAGCGCATAGATTGGTCAGAGAGACAGGTAAATATTATCCGCTTAGTTACAAAG 3461
|||
QY 42 CysThrValGluGluLysAla-----Leu 49
|||
Db 3462 GGGACTGTGGAGGAGAGATCATAGACGGCGGCAAAAGAGATGTTATTAGATCATCTG 3521
|||
QY 50 IleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGluAsn----- 64
|||
Db 3522 GTGATTCAGCGCATGGACACCACTGGCCGAGCATCTCGAAGAAACCACTCAGGAAGGTCC 3581
|||
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
|||
Db 3582 AACTCAATCCTTTTAAATAAGAGAGCTGACAGCTATTTTGAAATTTGGAGCAGGAGAT 3641
|||
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
|||
Db 3642 CTCCTTCAAGAACTGGAA-----GGGAGGAGGAATCA 3671
|||
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
|||
Db 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTGGGTGGCTGAA 3716
|||
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuAlaLysHisAlaGlnGlyThr 140
|||
Db 3717 ACAGAGAGAGATGAAGTGTCAACAGTGCACAGATGAAGACTTCTA---TCACAGTTTAA 3773
|||
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
|||
Db 3774 GTTGCCCACTTTGCCAATGCAAGATGAAGAAGAGCTA-----GAAGAGCCT 3821
|||
QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
|||
Db 3822 CCTCAACAGAGCTGGGATGATGATCATTT----- 3848
|||


```
Db 5148 AAGGCACCTGAAACAGCTGCACAAACCTGACAAAGGGCTCAACGTGCAGAACAGCTGGAA 5207
Qy 553 -----GluIleYsSerGluYsAsnGluYsLeuYsSer--- 564
Db 5208 CACACCCGAACTGCTGCTGAAATCGAGACCGGATAGCGAGTGCCTTAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGGAACCTATGGATTTTGTTCCAAGTTT 5327
Qy 565 -----LeuGluGlnMetHisAspValAlaYsIleYsLeuAla 577
Db 5328 ACAGAAATTTGATGCTCGAAATCGCAATAGTTATACAGATGGCTCATAGAAAAGGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleYsSerTrpAlaAlaYsLeuYsValCys 597
Db 5388 CAAGAAGAGAG-----GAGCAAAAGAGAAAGACGAC 5420
Qy 598 ValProIleGlnSerGlyAsnAsnYsHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGGTGAAGAAACCATTTTCGTCCAGAGCGCTCAGGCTCCAGCGG 5471
Qy 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492

RESULT 10
US-10-425-115-145108
; Sequence 145108, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 145108
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1
US-10-425-115-145108

Alignment Scores:
Pred. No.: 8,01e-07 Length: 661
Score: 161.00 Matches: 50
Percent Similarity: 44.51% Conservative: 23
Best Local Similarity: 30.49% Mismatches: 75
Query Match: 5.00% Indels: 16
DB: 20 Gaps: 5

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-425-115-145108 (1-661)
Qy 239 AspHisMetGlyAlaLeuGluSerProLYsValIleThrLeuGlnSerSerCysLYsSer 258
Db 6 GATGAAGTTGCTAAATCCAGCAATCACTGTCTCCACAAACAAACAAAGGTTAGGAA 65
Qy 259 SerGly-----ThrAspGlyThrLeuAspGlyAsnAspAlaPhe 271
Db 66 ATTCATTTTCAGAACTCTGAGATTCAGATGGGTGTCATCCTGATTTGATATGAT 125
Qy 272 GlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGln 291
Db 126 CATATGTGTTCTGTAATAATCAT---GACCTGCTTCTGMAATGATGACTCTCCAGTGCT 182
Qy 292 AspTrpGlyIleProAsp-----GluSerGlnArg-----ArgLeu 304
```

RESULT 11

```
US-10-369-493-26897
; Sequence 26897, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 26897
; LENGTH: 3045
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-26897
```

Alignment Scores:

```
Pred. No.: 1.04e-05 Length: 3045
Score: 160.00 Matches: 131
Percent Similarity: 37.00% Conservative: 111
Best Local Similarity: 20.03% Mismatches: 237
Query Match: 4.97% Indels: 176
DB: 17 Gaps: 29
```

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-369-493-26897 (1-3045)

```
Qy 17 AspValLYsHisValGluLYsIleLYsIleGluSerCysSerGluArgThrLYsIlePhe 36
Db 673 GACATCGAGCATTCGGCGGCAAGTGCAAACCGATCTCAAAAGAGAGGAGGATTCTCTATCT 732
Qy 37 ArgLeuTyrSerValCysThrValGluGluLYsAlaLeuIleAlaArgGlnAsnMet 56
Db 733 CGCTTGTG-----CAGGAAGAAGCT-----ACTCGCAAAACCGAG 765
Qy 57 ArgGlnAsnLYsAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrp 76
Db 766 CAGGCGGACCATTCGATTCGATTCCTTAACCCAGCTGAAC----- 807
Qy 77 GlyAlaSerTyrLeuPheAspLYsLeuAspHisPheHisSerSerGluThrProAspSer 96
Db 808 -----GATATCATGATACCGCGGAGCGCTGGCGCTTCTCTATCAA 849
```

QY 97 GlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerIleLeu 116
 DB 850 GGGGGGCCCCCATCTATGATGAGCGGCAACATCTTGAAGCTCTTCGATATATATG 909
 QY 117 SerSerIleGlyGlyGluGluAenGluValIleHisGluLeuGluGluAlaLys 135
 DB 910 AACTGTAC-----GAAGACACCGAGCAAGCTTGTTCAAACCGAGGAGCAAGGAC 963
 QY 136 HisAlaGlnGlyThrTySerSerAspSerThrPheGluHisGluAspHisIle 154
 DB 964 AAGGCAAGACAGCGGCGCAAGGAATCGCCACACGATCCCGCGAGTGAACAATCCAGAAG 1023
 QY 155 LeuSerAspGluGluSer-----ProAsn 162
 DB 1024 CTCAGCTCTGAACCAACCAAGCTTAAAGACACAGAGATGAGCAAAAGAGGCGGAC 1083
 QY 163 IlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyProSerAspThr 182
 DB 1084 CGCTTTTACCAAACTC-----AGAAAAAGGGATGGCTTGAGCCGACCGAGGAT 1134
 QY 183 Pro-----GlnArgAsnArgLysArgValGlnTyPheGluGly 195
 DB 1135 CCATCGCGCTCACTTGGCCCTCGACCAAGCTGTGAAAGACCAACCAAGATCGCTGGAG 1194
 QY 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysArg-----LysLysAla 214
 DB 1195 TCTGAAAGAATGCCCTGAACCAAGAAATCACTAATCTCAACAAATCGATCGAAGAGCC 1254
 QY 215 SerAspAspValThrAspProArgVal-----ThrAspProProValAspAsp 230
 DB 1255 AATGTGGCTCAAGACATGAAGCTTATCGAGTGGAGAGGATCTCTGGTGGATCTA 1314
 QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----ProLys 248
 DB 1315 CAACAACGTCAGNA-----GGGATATCAAAAGCTTGAACCAACAG 1356
 QY 249 ValIleThrLeuGlnSerSerCysLysSerGlyThrAspGlyThrLeuAspGlyAsn 268
 DB 1357 GTTCTGACCTCAAGCAAAATTTGGCGAGCAAGCGCCGCGCCGA-----TCT 1409
 QY 269 AspAlaPheGlyLeuTySerMetGlySerHisIleSerGlyLeuProGluAspMetLeu 288
 DB 1410 GAAGGATATTCAGTCTTGGATTTGGAGAACAAAGTCTTTGGAAGACCGAGCAACAGTGT 1469
 QY 289 AlaSerGlnAspTrp-----GlyLys 295
 DB 1470 GCAGCCGAAAGAGGCTTTCGACGACGAGCTTCTCAGACGAGCAAGGCTGGAAA 1529
 QY 296 IleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu 315
 DB 1530 GGTGGAGACGACCTCAAGAA-----CAGCGGCTCGCAAAAGATGATTT----- 1574
 QY 316 CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTy 335
 DB 1575 -----GGAGACTCAAAAGAAATGA 1592
 QY 336 ValIleGluAsn-HisArgIleTyGluGluProAlaThrPheGlnAlaPheGlnI 355
 DB 1593 GTGGTCCAGCAAGAAAGTCTCGAGGAGAG----- 1626
 QY 355 eAlaLeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLe 375
 DB 1627 -----ATCAGCAGGCTCAAGAAGGAGTTGACTTGGCCAGGAAGAATTCGCG 1673
 QY 375 u-----ValArgAlaAsnSerGluLeuAlaPheLysCysSerArgVa 389
 DB 1674 TGCGCCCAACATCGTCGAGATCAACCGGAACCCCAACATTCGCCAAAGCATCTCGCGT 1733
 QY 389 lGluVal-----AspTyTrileTy 395
 DB 1734 CGAACAACTCATGAGGAGCAAAACCTCACTACCGCTGACGCGCAGCTCTGTCCA 1793
 QY 395 rSerIleLeuSerCysMetLysSerLeu-----PheLeuGluHisThrGlnI 411

DB 1794 AGGACTGGTTCGCAAGTTCGAAGGACCGCGGAGGACACAAGGACTTGGATCCAGTGAGGA 1853
 QY 411 yLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuVa 431
 DB 1854 TGTTCATTCGTGAATGATCGGC-----CTTCT 1880
 QY 431 lAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn----- 446
 DB 1881 CAATCGTGTCTTGAAG---GAGCTCTCTTTGATGAGAAGACACAAGAGATGCTTTCAGAC 1937
 QY 447 -----ThrLysSerMetArgAsnSerSerGluAspGluGlu-----Cy 459
 DB 1938 TTTGCAAAATGAGTTCGAATGAGGCTCTGAGGAAGGCGGAGGACCGCAACAAGCGCG 1997
 QY 459 sMetThrGluLysArgCysSerHisTySerThrAlaThrArgAspIleGluLysThrI 479
 DB 1998 GCAGTTGCAGACATCATTCACAAATACGACGCGAAAGGAAAGGCTTGAAGAGGAGGT 2057
 QY 479 eSerGlyIleLysLysTyTrpLysGlnValGlnLysLeuValGlnGluHisGluI 499
 DB 2058 C-----AAGCTTCATAAAGACGAGATCAAGAAGTTGAAGCAGGAGATCACTGA 2105
 QY 499 uLysLysMetGluLeuLeuAsn-----MetTyAl 509
 DB 2106 GAAGAAACCTCACTCGCAACCAAGCAGCAGAGAGAGATATGCTCAAGAAAGCTACGA 2165
 QY 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaValIleArgIleTh 529
 DB 2166 AGAGCAAAATCAAAATCTCAATCGCGACCAACAAAAGGCTGCTCGGAGTGAAGTCAA 2225
 QY 529 rCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAspHisAs 546
 DB 2226 ACACCAAAACGAGCTGACGCGCTCAGAAAGGACGCGACTTGAAA-----GAAACCAA 2279
 QY 546 nTyGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuI 566
 DB 2280 CTTCTCGAAGATTAGACACACTCAGGCGACGAGATGAAGATGAGGAGGAAACAGGCTTCA 2339
 QY 566 uGlnMetHisAspValAlaLysLysLysLeuAlaGluAsp 579
 DB 2340 AGCGGATTAGCGCTGAGAAAGCGGAGGCTCAGCAAGGAT 2379
 RESULT 12
 US-10-956-157-859
 ; Sequence 859, Application US/10956157
 ; Publication No. US20050118625A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wyeth
 ; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
 ; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
 ; FILE REFERENCE: 031896-043000 (AM 101081)
 ; CURRENT APPLICATION NUMBER: US/10/956,157
 ; NUMBER OF SEQ ID NOS: 319805
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 859
 ; LENGTH: 5883
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-956-157-859
 Alignment Scores:
 Pred. No.: 4,61e-05 Length: 5883
 Score: 158.00 Matches: 131
 Percent Similarity: 35.78% Conservative: 98
 Best Local Similarity: 20.47% Mismatches: 207
 Query Match: 4.91% Indels: 205
 DB: 21 Gaps: 28
 SEQ3_D1219_COPY_650_1270 (1-621) x US-10-956-157-859 (1-5883)

Qy	22	GlulysllelyslleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal	41
Db	2878	CAAGAAGCTGCAGCTGGAGAAGGTGACCCACGAGCGGAAGCTGAAAAAGCTG	2928
Qy	42	CysThrValGluGluLysAlaLeuLleLysLeuAlaArgGlnAsnMetArgGlnAsnLysAla	61
Db	2929	-----GAGGAGGAGCAGACTCATCTCGAGGACCAGAACTGCAAGCTGGCCCAAG	2976
Qy	62	ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu	81
Db	2976	-----	2976
Qy	82	PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu	101
Db	2977	-----GAA	2979
Qy	102	GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly	121
Db	2980	AAGAAACTGTGGAAGACAGAANTAGCTGAGTTCCACCAACCTTCACAGAA	3030
Qy	122	GlulGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyr	141
Db	3031	GAGGAGAGAACTTAAGACCTCGCCAAAGCTCAAGAAACAGCAT	3075
Qy	142	SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerPr	161
Db	3076	-----GAGGCAATGATCACTGACTTGGAGAGCGCTCCGACGGAGGAGAACGAC	3126
Qy	161	oAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAs	181
Db	3127	CGACAGAGCTGGGAGAGACCGCGCGGAAGC	3165
Qy	181	pThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLys	201
Db	3166	GACTCCACAGACCTCAGCGACCA	3210
Qy	201	sThrGlyAspGlyGlyAsnAlaLysLysArgLysAlaSerAspValThrAspPr	221
Db	3211	ATCGCGGAGCTCAAGATGCAGCTGGCCAGAAAGAGG	3247
Qy	221	oArgValThrAspProProValAspAspAspGluArgLys	234
Db	3248	-AGAGCTCCAGGCGCGCTGGCCAGAGTGGAGAGGAAGTGCCCAAGAAACATGGCC	3306
Qy	235	-----AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal	250
Db	3307	CTCAAGAAGATCCGGGAGCTGGAAATCTCAGACTCTGAATCTCCAGGAAGACCTGGAGTCT	3366
Qy	250	eThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAl	270
Db	3367	GAGCGTCTCCAGGATTAAGCTGAGAACGACGAAACGGGACCTTGGGG	3415
Qy	270	aPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSe	290
Db	3416	-----AAGAGCTAGAGGCTCTGAAA	3435
Qy	290	rGlnAspTrpGlyLys	306
Db	3436	ACAGAGTTGAGGACACGCTGGATTCACAGCT-GCCACAGCAGGAGCTCAGGTCAAAACG	3494
Qy	307	-----ValLeuLysProLysMet	316
Db	3495	TGACGAGGAGGTGAACATCTTGAGAGAGACCTTGAGGAGGAGGCCAAGACCCACGAGGC	3554
Qy	316	sGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVa	336
Db	3555	CCAGATCCAGGAGATGAGGCAGAGCACTCAACAGCCGTGGAGGAGCTGGCGGAGCAGCT	3614
Qy	336	lIleGluAsnHisArgIle	353
Db	3615	GGAGCAGCAAGACCGGTGAAGCAAACTCCAGAAAGCAAGACAGACTCTGGAGAACGA	3674
Qy	353	eGlnIleAlaLeuSerTrpIleAlaAlaLeuLeuVal	371

RESULT 13

RESUL 13
US-10-696-909A-48

US-10-030-202A-48
: Sequence 48. Application US/10696909A; sequence 48, Application 00, 100
; Publication No. US20050118604A1; PUBLICATION NO. US20
: GENERAL INFORMATION:

APPLICANT: LORENS, JAMES B.

APPLICANT: LORENZ, JAMES B.
APPLICANT: ATCHISON, ROBERT E.

APPLICANT: Atchison, Robert
APPLICANT: Frier, Anabelle

APPLICANT: FRIERA, ANABELLA
APPLICANT: HOLLAND, SACHA

APPLICANT: Piceal Pharmaceuticals Inc

APPLICANT: Rigel Pharmaceuticals, Inc.

; TITLE OF INVENTION: Modulators of Angiogenesis
; FIVE REFERENCE: 031044-005820US

FILE REFERENCE: 021044-005820US

; CURRENT APPLICATION NUMBER: US/1

; CURRENT FILING DATE: 2003-10-29

; PRIOR APPLICATION NUMBER: US 60/512,251

; PRIOR FILING DATE: 2003-10-17

384	Qy	eLysCysSerArg-----ValGluValAspTyrIleTyrSerIleLeuSerCysMetLy	402
3958	Db	CAAGTCAACAAGCTGCAGGTGGAGCTGGACACAGTACCAGGCTTCTTCAGCAGTCCGA	4017
402	Qy	sSerLeuPhe-----LeuGluHisIth	409
4018	Db	CAGCAAGTCCAGCAAGCTCAACCAAGGACTTCTCGCGCTGGAGTCCAGCTGCAGGACAC	4077
409	Qy	rGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLy	429
4078	Db	TCAGGAGCTGCTGCAGGAG-----GAGAACCGCAGAGTGCAGCTGCAGCACAA	4128
429	Qy	sLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsnThrLysSe	449
4129	Db	GCCTC-----AAGCAGGTGGAGGACGAGAAG-----AATTC	4158
449	Qy	rMetArgAsnSerSerGluAspGluGluCysMet-----ThrGluLysArgCysSerHi	467
4159	Db	CTTCGGGAGCAGCTGGAGGAGGAGGAGGCCAAGCACAACTGGAGAAGCAGATCGCCAC	4218
467	Qy	sYrSerThrAlaThrArgAspIleGluLys-----	477
4219	Db	CCTCATGCCCGCTGGTGGCCGACATCAAAAAGAGATGGAGACAGTGTGGGGTGCCTGGA	4278
478	Qy	ThrIleSerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHi	497
4279	Db	AACTGCTGAGGAGTGAAGAGGAGAGCTCCAGAAGGACCTGGAGGGCTTGAGCCAGCGGCA	4338
497	Qy	sGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluTh	517
4339	Db	CGAGGAGAAG-----GTGGCGCGCTAC--GACAAGCTGGAGAAGACCAAGAC	4383
517	Qy	rSerLysSerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnVa	537
4384	Db	G-----CGGCTGCAGCAGGAGCT	4401
537	Qy	lGlyAspLeuLysLeu--LeuAspHis-----AsnTy	547
4402	Db	GGAGCACTCTGCTGGAGCTTGACACCAAGCCGAGAGCGGTGCNACTGGAGAAGAA	4461
547	Qy	rGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGl	567
4462	Db	GCAGAAGAAGTTTGACCAGCTCTCGCGGAGGAGAGACCATCTCTCCCAAGTATGCAGA	4521
567	Qy	nMetHisAspValAlaLysLysLeuAlaGluAspGlu	580
4522	Db	GGAGCGCGACCGGGCTGACGGGAGGAGCCCGAGAGAGGAG	4561

RESULT 15

US-10-107-782-35
; Sequence 35, Application US/10107782
; Publication No. US20040018970A1

: GENERAL INFORMATION:

APPLICANT: Boldoq, Ferenc,

; APPLICANT: Casman, Stacie

APPLICANT: Colman, Steve,

APPLICANT: Edinger, Shlomit,

APPLICANT: Gangolli, Esha

APPLICANT: Kekuda, Ramesh

; APPLICANT: Li, Li,

APPLICANT: Liu, Xiaohong,

; APPLICANT: Malyankar, Uriel,

APPLICANT: Miller, Charle

; APPLICANT: Millet, Isabel

; APPLICANT: Patturajan, Meera,

APPLICANT: Rothenberg, Mark,
APPLICANT: Rothenberg, Mark,
APPLICANT: Rothenberg, Mark,

APPLICANT: SCIORE, PAUL, CHRISTOPHER

APPLICANT: Shenoy, Suresh

APPLICANT: Shimkets, Richard,
si Tinghong

APPLICANT: SI, Jingsheng,
APPLICANT: Smithson Glen

```

; APPLICANT: SMITHSON, GlennDa,
: : savytek kimbber]v
: APPLICANT:

```

APPLICANT: SPYTEK, KIMBERLY,

```

; APPLICANT: Stone, David.
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 35
; LENGTH: 7396
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)..(6016)
US-10-107-782-35

Alignment Scores:
Pred. No.: 6,53e-05 Length: 7396
Score: 158.00 Matches: 133
Percent Similarity: 36.44% Conservative: 98
Best Local Similarity: 20.98% Mismatches: 210
Query Match: 4.91% Indels: 194
DB: 17 Gaps: 29

SEQ3_D1219_COPY_650_1270 (1-621) x US-10-107-782-35 (1-7396)

Qy 22 GluLysIleLeuValIleHisGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 3017 CAGAAGCTCGACGTGGGAAGGTGACACCGACGCGGCAAGCTGAAAGAGCTG----- 3067
Qy 42 CysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAla 61
Db 3068 -----GAGGAGGAGCAGATCATCTCTGGAGGACCAAGAACTGCAGCTGCCAAG--- 3115
Qy 62 ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
Db 3115 ----- 3115
Qy 82 PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101
Db 3116 -----GAA 3118
Qy 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
Db 3119 AAGAAACTGCTGGAGACAGAAATAGCTGAGTTCCACCACCACTCCACAGAA----- 3169
Qy 122 GluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
Db 3170 GAGGAGGAGAAATCTAAGAGCTCCCAAGCTCAAGACAGCAT----- 3214
Qy 142 SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGlu-SerPr 161
Db 3215 -----GAGGCAATGATCACTGACTTTGGAAGAGCGCCTCCGCGAGGAGGAGGAGCAG 3265

```

Qy	161	oAenillePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAs	181
Db	3266	CGACAGGAGCTGGAGAAAGACCGCGCGAAGC-----TGGAGG-----GA	3304
Qy	181	pThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLy	201
Db	3305	GACTCCACAGACTTCACGCCACCA-----TCGCCGAGCTCCAGGCCACG	3349
Qy	201	sThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspPr	221
Db	3350	ATCGCGAGCTCAAGATGCGAGCTGGCCAGAAAGAGG-----3386	
Qy	221	oArgValThrAspProValAspAspAspGluArgLys-----234	
Db	3387	AGAGGCTCCAGCGCGCCCTGGCCAGAGTTCGAAGAGGAAGCTGCCAAGAAGACATGGCC	3445
Qy	235	-----AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal-----11	250
Db	3446	CTCAAGAAGATCCCGGAGCTCGAATCTCAGATCTCTGAATCCAGGAAGACCTCGAGTCT	3505
Qy	250	eThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAl	270
Db	3506	GAGGCTCTCCAGGAATAAGCTGACGAACGACGAACCGGACCTTGGGAA-GAG-----3559	
Qy	270	aPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSe	290
Db	3560	-----CTAGAGCGCTGAAAAACAGAGTTGGAGGACACGCTGTGATTCACAGCTGCCCA	3612
Qy	290	rGlnAspTrpGlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysPr	310
Db	3613	GCAGGAGCTCAGGTCAAAACGTGACGAGGAGGTGAACATCTCAAGAAGACCTCGAGGA	3672
Qy	310	oLysMetAlaLysLeu-----CysGlnValLeuHisLeuSerAspAlaCysThrSerMe	328
Db	3673	GGAG--GCCAAGACCCACGAGCGCCAGATCCAGGAGATGAGGAGAGCACTCACAGGC	3729
Qy	328	tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIle-----TyrGluGl	345
Db	3730	CGTGGAGGAGCTGCGGAGCAGCTGGAGACAGACGAAGCGGTGAAGCAACCTCGAGAA	3789
Qy	345	uProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuVal	365
Db	3790	GGCAACGACACTCTGGAGAACGAGCGGGGGAGCTG-----GCCAAGAGGT	3837
Qy	365	lLysGlnIleLeu-----SerHisLysGluSerLeuVal-----376	
Db	3838	GAAGTCTGCTGACGGCGGAAGGGACTCGGAGCACAAAGCAAGAAAGTGGAGGCGCA	3897
Qy	377	-----ArgAlaAsnSerGluLeuAlaPh	384
Db	3898	GCTGCAGGAGCTGCAGTCAAGTTCAACGAGGGAGACCGGTGCGCACAGAGTGGCCGA	3957
Qy	384	eLysCysSerArg-----ValGluValAspTyrIleTyrSerIleLeuSerCysMetLy	402
Db	3958	CAAGTCCACCAAGCTGCAGGTGGAGCTGGAACAAGTACCGGGGTCTTCAGCCAGTCCGA	4017
Qy	402	sSerLeuPhe-----LeuGluHisTh	409
Db	4018	CAGCAAGTCCAGCAAGCTCAACAGGACTTCTCCGCGCTCGAGTCCAGCTGCAGGACAC	4077
Qy	409	rGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLy	429
Db	4078	TCAGGAGCTGCTGCAGGAG-----GAGAACCAGCAAGCTGAGCTGAGCACCAA	4128
Qy	429	sLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsnThrLysSe	449
Db	4129	GCTC-----AAGCAGGTGGAGACGAGAG-----AATTC	4158
Qy	449	rMetArgAsnSerSerGluAspGluGluCysMet-----ThrGluLysArgCysSerHi	467
Db	4159	CTTCCGGGAGCAGCTGGAGGAGGAGGCGCAACCAACCTGGAAGAGCAGATCGCCAC	4218

Db 1183 QRRHLTVLKPQAKLCOVLHLSDACTSMVGNFLEYVYIENHRIYEEPATTFQAFQIALSWI 1242
QY 361 AALLVKQILSHKESVRANSELAFKCSRVEDYIYISILSCMSKSLFLEHTQGLQFQFCFGTN 420
Db 1243 AALLVKQILSHKESVRANSELAFKCSRVEDYIYISILSCMSKSLFLEHTQGLQFQFCFGTN 1302
QY 421 SKOSVVSTKL VNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTTIS 480
Db 1303 SKOSVVSTKL VNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTTIS 1362
QY 481 GIKKKYKQVOKLVQEHKQKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGL 540
Db 1363 GIKKKYKQVOKLVQEHKQKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGL 1422
QY 541 KLLDHNRYERKFEI EKNECLKSLEQMDHVAKKLAEDBACWINRIKSWAAKLKVCVPI 600
Db 1423 KLLDHNRYERKFEI EKNECLKSLEQMDHVAKKLAEDBACWINRIKSWAAKLKVCVPI 1482
QY 601 QSGNNKHFSGSSNISQAPDV 621
Db 1483 QSGNNKHFSGSSNISQAPDV 1503

RESULT 2
T23056
chromodomain helicase H06001.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23056
R:Barlow, K.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19665
A:Accession: T23056
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1465 <WIL>
A:Cross-references: UNIPROT:O17909; EMBL:Z92970; PIDN:CAB07481.1; GSPDB:GN00019; CESP:HC
A:Experimental source: clone H06001
C:Genetics:
A:Gene: CESP:H06001.2
A:Map position: 1
A:Introns: 44/1; 91/3; 170/3; 377/3; 494/3; 1046/2; 1099/3; 1242/3; 1298/1
C:Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 5.1%; Score 163.5; DB 2; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.058;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADATILFGSSLPNSHDVHKVEKIKIESCSERTKIPRLYSVCTVVEKALILARQNMQRNKA 61
Db 799 ADTVIIFDSNDWNPQNDLQAMSRHRIGQTKTNIYRLVTKGSVEE-----BI 845
QY 62 VENLNRSLTHALLMWCASYLFDKLDH-----FHSETPDSCGVSEQSTMDGV 108
Db 846 VERARKLV-----LDHLVTRQMDTTKTVLSKNATASGVDPDKQLSAI 891
QY 109 IHEFGSI-----LSKGGGENEV-----KCLLLAKHAQGYTSDSTIFGB-DHKLSDERS 160
Db 892 L-KFGAVELPKEGEQEPEVDIDRLMCAETREAEVEEVMKENELSSFKYANFAIDEE 950
QY 161 PNIF-----WSKLLGKNPMWKYPSTPQNRKRV-----Q 191
Db 951 KOIAATADWAAII-----PEEDNRILIEERMKELAEMLNAPRQKQPIQ 997
QY 192 YEGSEA-SPKTDGCGNAKKRKASDDVTDPRV-----TDPVDDERK 234
Db 998 VVEDDDGDDDEBDDTGKXXXXKAVGNFTIPEIKRFRKFRKMPNLNLEIAQAELE 1057
QY 235 ASGKHMGALSPKVITLQSSCKSGTGDGLDGNDAFGLYSMGSHISGIPEDMLASQDWG 294
Db 1058 EHSTDMMKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
QY 295 KIPDESQ-----RRLHVLTKPKQAKLCOVLHLSDACTSMVGNFLEYV 336

Db 1093 AAESKKDIERKFKPHTCDVNLKQIERSHAELKP-----LHEILKSEETKTS----- 1139
QY 337 IENHRIYEEPATTF-----FQAFQIAL-----SWIAALIVKQI-LGKHESL 375
Db 1140 -----FKPPANAKLQKGDVWDWSPDDSSALLLGVWKYGYGWEAIKMDPTLGLADKPIFI 1193
QY 376 VRANSELAFKCSRVEDYIYISILSCMSKSLFLEHTQGLQFQFCFGTNKOSVSTKLVNLSL 435
Db 1194 KDKTKKPGQKNLQVRVDYLLKLM-----SKDKVKVTE----- 1225
QY 436 SGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTTISGIIKKYKQVOKLVQ 495
Db 1226 -----KKERKRKADDPVGPPEKKCGKRTNNVPOEEK-----KCKEKK----- 1262
QY 496 EHEEKQKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGLKLLDHNRYERKFDIEK 555
Db 1263 --EKNSSSLK---DQLALLSIDKSLYGALEDSAKPFLECVKLCMPVHKYTKKLEKQ 1317
QY 556 SEKNECLKS-----LEQMDHVAKKLAEDBACWINRIKSWAAKLKVCVP 599
Db 1318 EAKNQADEAKVLTRLGDSFLENLETLIKKPKTNIRKWNVYLWIFLCKFTLREP 1371

RESULT 3
C84507
hypothetical protein At2g13370 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84507
R:Lilin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
eius, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84507
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1738 <STO>
A:Cross-references: UNIPROT:Q9SI41; GB:AE002093; NID:g4733988; PIDN:AAD28668.1; GSPDB:GN
C:Genetics:
A:Gene: At2g13370
A:Map position: 2
C:Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 5.0%; Score 160.5; DB 2; Length 1738;
Best Local Similarity 18.5%; Pred. No. 0.11;
Matches 138; Conservative 104; Mismatches 234; Indels 269; Gaps 31;

QY 2 ADATILFGSSLPNSHDVHKVEKIKIESCSERTKIPRLYSVCTVVEKALILARQNM----- 56
Db 1041 ADTVIIFDSNDWNPQNDLQAMSRHRIGQGVVNIYRFTVTSKVEBEILERAKRMVLDHL 1100
QY 57 -----RQKAVENLRSLTHALLMWCASYLFDKLDHSHSETSPDSGVSEFS 103
Db 1101 VTQKLNABERLEKRETKKSGNFDKNLSAILRFGAEELFKE----- 1141
QY 104 IMDGVIHEFSSILSSKGGGENEVKLCLE-----LEAKHAQGYTSSDSTLFGEDH 152
Db 1142 -----DKNDEESKRLLSMDIDBILERAQVEEKH---TDTEHELLGAPK 1184
QY 153 I-KLSDSESPNIFWSKLLGKNPMWKYPSTPQNRKRVQYPEGSEASPKTDGCGNAKKR 211
Db 1185 VANFCNAEDDGGCSFWSRWIKPDSVVTAEBALAPRAARNTKSYVDPSHPD-----RTSKRK 1238
QY 212 KIASDDVTDPRVTDPPVDDDERKASGKHMGALSPKVITLQSSCKSGTGDGLDGNDAF 271
Db 1239 KKGSE-----PPEHTERSQRRRTKTEYFVPTPLLEGTSAQVRGWSYGNLPKRDAQ 1288
QY 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHVLTKPKMAKLCOV 318
Db 1289 RPYRTVMKFGNHNQACIAEEVGGVVE-----AAPEEAQVELFDA----- 1328

QY 319 LHLSDACTSV--GNF-----LKVYIENHRIYEPATTFQ 351
Db 1329 --LIDGCKSVETGCTNFKPGVLDFFGVPVKANELLKRVQGLQLLSKRISRYNDPISQFR 1386
QY 352 AFQIALSWIA-----ALLVKQILSH-----KESLVANSELAF--KCSRVEV 391
Db 1387 -----VLSYKPSNWSKCGNQIDDARLLGLIYHGFNGWKEKIRLDESGLTKKIAPVEL 1442
QY 392 DYIYSILSKMSLFLEHTQLOPDCF-----GTNSQSVVSTKLVNESLSG---ATVRDE 443
Db 1443 QHETFLPRAPNLKERTALLENELAAAGCKNTAKASRNSKKVKNLQINQKAPADR 1502
QY 444 K-----INTK-----SNRNSSE-----DEE 458
Db 1503 RGKSGPANVLLSTKDGPRKTOKAEPLVKEEGEMSDGEVYEQKEQKMWENCEDVLADE 1562
QY 459 CMTEKCSHYSTATRDI--EKTISGKKYK---KQVKLVQSH--KMWELN 506
Db 1563 IKTGLRLQRLQTSADLPKEKJFKIRRYLEILGRIDAIVLEHEEDLYKQDRMTWRLWN 1622
QY 507 -----MYADKKQKLETSKVSVEAAVIRITCSRTSTQVGDLLKLLDHNVERKF 551
Db 1623 YVSTFNSLSDRLNQIYSKLEKEEBEGVGPS--HLNGSR-----NFRQ--1666
QY 552 DEIKSEKNECKLSLEQMH-----DVAK 573
Db 1667 QKPKTAGNS--QGSQVHKGIDTAK 1689

RESULT 4

H89567
protein T08A9.1 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: H89567
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gac/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: H89567
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1330 <STO>
A:Cross-references: UNIPROT:Q22342; GB:chr_X; PIDN:AAA81420.1; PID:G1065501; GSPDB:GN000
C:Genetics:
A:Gene: T08A9.1
A:Map position: X

Query Match 4.9%; Score 156.5; DB 2; Length 1330;
Best Local Similarity 19.1%; Pred. No. 0.14;
Matches 133; Conservative 111; Mismatches 225; Indels 227; Gaps 33;
QY 58 QNKAVENTLNRSL-----THALLMWGASYLFDKLDHF-----HSETPDGSGVSPQSIM 105
Db 446 EESREQLGRKKHFLHALF-----PGLFDNLPEFYVYKAPLEKYDTDLPIAKEYIKDLR 501
QY 106 DGV-----IHFSSILSSKGGENEVKLLLEAKHAQCTYSSSTLFGEDHKK- 154
Db 502 DALPELEPFLKVLTPNVASKLATKSKQH-----PSFGQTRVES 540
QY 155 -LSDEE-----SPNIFWKLJGGKNP-----MWKYPSTPQNRKR-----VQ 191
Db 541 FLTDEPMRMARSHFNYSPPAALWSDGDSFVQPLMCRSPENALRESQCRSIPYVPSLQ 600
QY 192 YFGSEASPKTGDGNAKTKKASDDVTDPRVTDPPVD--DDBKASGKD--HMGAL---244
Db 601 QLEGLD-GPAPG-----SSAPISIENTSSRINFKQSGRQSQSSQDLHHVGSVSS 649
QY 245 -----ESP-----KVITLQSSCKSGTGTGLDGNDAF-GLYSMSGSHISGIPED-MLAS 290

Db 650 DTSLLGHETPVKMDTVLMTENKSELNLSNMNDSNDSIDSLLDVYEHIDYSIEDILLGL 709
QY 291 QDWGKIPDESQRRLHTV-----LKPRMAKLQVHLHS-----DACTSMVGNF 332
Db 710 SDHTVQLEDAVKLSKMQNIVANMTPRNEKLFYLLHQTAPRDFEIIHDGDKLKNRVKEL 769
QY 333 -----LEVVIENHRIYEPATTFQFOIALSWIAALLVKQILSHKESLVANSELAPFC 386
Db 770 GESVESLETELENKST-----SNLEAKLAEMTVSHKKEIENTQAE-----811
QY 387 SRVEDVIYSILSKMSLFLEHTQLOPDCF-----GTNSQSVVSTKLVNESLSGATVRDESKIN 446
Db 812 -----CIRMSVEFE--LWTSMSRQSKQESKQREIEELKAKLEKQNIH 856
QY 447 TKSMRNSDEDECMTEKRCSHYSTATRDIEKTIISGKKYKQVQKLVQ-----495
Db 857 EKALRNDPYSEE-----YKRTLTAER--AELEKPFQRIEVTITKAVECKKDEAFA 905
QY 496 -----EHEEKMWELNMYADKKQKLET--SKSVEAAVIT--RITCS 531
Db 906 ROEKTLEIENRVLSENESKSKLEAMNR---EKEQLESIRQMEGEVILDEFNALKES 962
QY 532 RTSTQVGDLL-----KLADHNYERKFDEI--KSEK--NECLKSLEQMHDAVAKK 575
Db 963 QPSAQIKDRPAIRSRMERSTDGKCLDTNKLVDLIDNLKVLKQLVNEREQIRDVILKSQ 1022
QY 576 LAEDEACWINRIKSWAAKLVKVCPIQSGNNKHFSGS 611
Db 1023 PGGDQV--LQFLENSPTPTVVDMEAYFNELTNFGS 1056

RESULT 5

T46337
hypothetical protein DKFZp434O2413.1 - human (fragment)
C:Species: Homo sapiens (nan)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46337
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Accession: T46337
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-992 <AAA>
A:Cross-references: UNIPROT:Q9NTH6; EMBL:AL137265
A:Experimental source: adult testis; clone DKFZp434O2413
C:Genetics:
A:Note: DKFZp434O2413.1

Query Match 4.6%; Score 147.5; DB 2; Length 992;
Best Local Similarity 19.0%; Pred. No. 0.33;
Matches 125; Conservative 98; Mismatches 219; Indels 215; Gaps 26;
QY 122 RENEVKLLLEAKHAQCTYSSDSTLFGEDHKKLSDESPNIFWKLGG-----171
Db 91 EKSEPKICRNLVTPKADPT-----GSEPAKASEKPAE---DVTADAGEEGSRREAA 139
QY 172 KNPMMKY-----PSDTPQ-----RNRKRVQVFEGSEASPKTGDGNAKGRKKASDDVTD 220
Db 140 KFPKKAASALEGSSDASQELISEHMKEPQLSDSTASDPKSPHGLDFGRSRISEHLLD 199
QY 221 PRVTD-----PVDDDERKASGKHMGAL--PKVITLQSSC 256
Db 200 VDLSPVLGACRQAQQLGIEDKDSQSSQDELQSKQSGLEERLSPPLPHEERAQSP 259
QY 257 KS-----SGTDGTLN-----GNDAFLYSM-----276
Db 260 RSLATEEPFQPGEPQFEWKEAELGEDSAAASLSLQSLQREQOAPSPPAACEKGEHQSHQ 319
QY 277 -----GSHISGIPEDMLA-----SODWGIKIPDE 299
Db 320 AELGPGQEAEDPEKVAVSPPTPPVPSVPRSTEPVAPPEQLSEALKAMEEVAQVLEQ 379

**
**
**

**
**
**

**

**
**
**
**

**
**
**

Db 1344 ERNSQEQLEDEWEAKQNLHERHISTINIQLSDSKKKLQDPFASTVESLEEGKKRFQKEIIS 1403
QY 493 LVQHEEKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDKL-LDH----- 545
Db 1404 LTOQVEEK-----AAAYDKLETK-----NRLQQLDLDLVLDNQRLVS 1444
QY 546 ---NYRKFDEIKSEKNECKLSLEQMHVAKKLADE 580
Db 1445 NLEKKQKQFDQLAEKKNISKYADRRDRAEAAREKE 1482
RESULT 10
S39082
myosin heavy chain, embryonic - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 03-Feb-1994 #sequence revision 03-Feb-1994 #text_change 09-Jul-2004
C:Accession: S39082; S24349; A30170; S01265
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
submitted to the EMBL Data Library, August 1991
A:Description: Analysis of the chicken fast myosin heavy chain family: Localization of
A:Reference number: S39081
A:Accession: S39082
A:Molecule type: mRNA
A:Residues: 1-741 <MOO>
A:Cross-references: UNIPROT:091983; EMBL:M74086
A:Experimental source: clone Cemb3
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24349
A:Molecule type: mRNA
A:Residues: 1-12, 14-741 <MOO2>
A:Cross-references: EMBL:M74086
A:Experimental source: clone Cemb3
R:LaGrutta, A.A.; McCarthy, J.G.; Scherzinger, C.A.; Heywood, S.M.
DNA 8, 39-50, 1989
A:Title: Identification and developmental expression of a novel embryonic myosin heavy-c
A:Reference number: A30170; MUID:89210285; PMID:2707122
A:Accession: A30170
A:Molecule type: DNA
A:Residues: 723-741 <LAG>
A:Cross-references: GB:M24691; NID:G341219; PIDN:AAA48950.1; PID:G531186
R:McCarthy, J.G.; Heywood, S.M.
Nucleic Acids Res. 15, 8069-8085, 1987
A:Title: A long polypyrimidine/polypurine tract induces an altered DNA conformation on t
A:Reference number: S01265; MUID:88040428; PMID:3671071
A:Accession: S01265
A:Molecule type: DNA
A:Residues: 723-741 <MCC>
A:Cross-references: EMBL:X06251; NID:G63600; PIDN:CAA29593.1; PID:G63601
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; muscle
Query Match 4.4%; Score 142; DB 2; Length 741;
Best Local Similarity 19.7%; Pred. No. 0.49;
Matches 122; Conservative 99; Mismatches 214; Indels 184; Gaps 30;
QY 50 ILARQNMQRKAVENLRSLTHA---LLMWGASYLFDKLDHFSSTPDSGVSPFQSI 106
Db 147 LLREQYEEQEAQKGLQALSKANSVAQWRKTYETDAIQ--RTELESAKKKLAQLQD 204
QY 107 GVTH-----EFSILSSKGGENEVK-----LCLLEAKH-----AQGTYS 142
Db 205 AEEHVAENVAKCASLETKQRLQNEVEDLMIDVERANACALDKKQKFDKILAEWKQK 264
QY 143 SDSLTFGEHIKLSDESPNIFWSKLLGGNPMWKYPS-----DTPQRNKRVRQYPEGSE 197
Db 265 YEET---QAELEASQKESRSLT-TELFKMKV---AYEESLDHLQTLKRENKNLQ-Q-EISD 316
QY 198 ASPKTCGDGNA-----KKPKASDDVTDPRV-----TDPVDDDE----- 232
Db 317 LTEQIAEGGKATHELEKVKKQIEQEKSEIQALEEAEASLEEEGKILRLQLELNQVKE 376

QY 233 --RKASGKD-----HMGALSPKVVITLQSSCKSSGTDTGLDG-----NDAPGLYSM 276
Db 377 IDRKIAEKDEEDQLKRNHLRIVES-----LQSS-----LDAETSRNEALRL----- 419
QY 277 GSHISGIPEDM-LASQDWGKIPDESORRL---HTVLKPKMAKLCQVHLSDACTSMVGNF 332
Db 420 KKKMEGDLNEMBIQLSHANRVAAEAQKRLNRTQAVLKDTQ-----IHLDDALRTQ----- 469
QY 333 LEYVJENHRIVEEPATTFQAFQIALSWIAALLVKQLSHKESLVRANSELAFKCSRVEVD 392
Db 470 -EVLKEQVAMVERRANLQA---EIEELRAALEQTERKKV---AEQLMDASERVQ--- 519
QY 393 YIYSILSCWKSILFLEHTQQLQDFCGTNSKQSVSTKLVNESLSGATVRDEKINTKSMWN 452
Db 520 -----LLHTQNTSL-----INTKKLETDLAQIQSEMEDTQIEARNTSEKAKK 562
QY 453 SSEDCECTEK-----RCSHYSTATRIEKTISGIK-----KKYKKQVQKLVQ 495
Db 563 AITDAMMAEELKKQDTSALHRMKNKLDQTVKQLHRLDRAEQALKGGKKQIKLEA 622
QY 496 EHEEKKMELLNMYADKKLETSKSVAAVIRITCSRTSTQVGDKLKLDHNYERKFDEIK 555
Db 623 RVRELEGE---VDAEQKSAEAVKV-----RKYERRVKELT 656
QY 556 SEKNECKLSLEQMHVAKK 574
Db 657 YQSEEDLKNILRLQLDVK 675
RESULT 11
T02572
hypothetical protein At2g39300 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T16B24.6
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02572; F84815
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02572
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-768 <ROU>
A:Cross-references: UNIPROT:080951; EMBL:AC004697; NID:G3402671; PID:G3402677
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84815
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-768 <STO>
A:Cross-references: GB:AE002093; NID:G3402677; PIDN:AAC28980.1; GSPDB:GN00139
C:Genetics:
A:Gene: T16B24.6; At2g39300
A:Map position: 2
A:Introns: 80/2; 665/3
Query Match 4.4%; Score 142; DB 2; Length 768;
Best Local Similarity 19.2%; Pred. No. 0.51;
Matches 107; Conservative 94; Mismatches 201; Indels 154; Gaps 25;
QY 84 KLDHFHSSSTPDSGVSPFQSIIMDGVHFEPSILSS-----KGEENEVKL-----CLLL 132
Db 25 KDDHFQRLVNSPKGLNKSQSEVGAALRRSRSLSSAAAFVIGDTSNQHRLNHSRCLTP 84
QY 133 EAKHAQGTYSST-----LFGEDHIKLSDESPNIFWSKLLGGNPMWKYP 179

Qy	430	LVNESLSGATVRDEKINTKSNRNSSEDE	---CWTEKRCSHYSTATPDIK	-----	477
Db	1331	L-----KQVEDEK	---NSFREOLESEEEAKNLEKQATLHAQVADMKQEDSVCC	1380	
Qy	478	--TISGIKKKYKQVQVKLVQHESEKKQWELNMYADKKQKLETSKSVAAVIRITCSRTST	535		
Db	1381	LETAEEVKRKLQDKLEGLSQRHEEK	---VAAAY-DKLEKTKT-	1422	
Qy	536	QVGDKL-LDH-----	NYERKPFDEIKSEKNECLKSLQPMHDVAKKLADE	580	
Db	1422	ELDDLLVDLHORASCNLEKKQKFFOOLAAEKTISAKYAEERDRABAEAREKE	1476		

Search completed: July 12, 2005, 01:03:36
Job time : 17.6667 secs

A;Cross-references: GB:M31013; NID:gl89035; PIDN:AAA36349.1; PTD:gl89036
R;Toothaker, L.E.; Gonzales, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.;
Blood 78, 1826-1833, 1991
A;Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' CDNA clones, C
A;Reference number: 152562; MUID:92003925; PMID:1912569
A;Accession: I52562
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-52, 'EAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>
A;Cross-references: GB:M81105; NID:gl88988; PIDN:AAA59888.1; PTD:G553596
R;Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A;Title: Identification and overlapping expression of multiple unconventional myosin gen
A;Reference number: A55758; MUID:94294418; PMID:8022818
A;Accession: I61692
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 182-218 <BM>
A;Cross-references: GB:L29141; NID:G457249; PIDN:AAA20904.1; PTD:G531134
C;Genetics:
A;Gene: GDB:MYH9
A;Cross-references: GDB:120216; OMIM:160775
A;Map position: 22q12.3-22q13.1
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP: coiled coil; hydrolase; methylated amino acid; nucleotid
F;84-764/Domain: myosin motor domain homology <MMOT>
F;174-181/Region: nucleotide-binding motif A (p-loop)
F;552-565/Region: actin binding #status predicted
F;626-640/Region: actin binding #status predicted
F;837-1938/Domain: coiled coil #status predicted <COL>
F;837-1277/Domain: S2 #status predicted <DS2>
F;1278-1961/Domain: light meromyosin #status predicted <LMM>
F;1939-1961/Domain: carboxyl-terminal <CBT>
F;125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;180/Binding site: ATP (Lys) #status predicted
F;694.704/Active site: Cys #status predicted

Query Match	4.4%;	Score 141;	DB 1;	Length 1961;
Best Local Similarity	19.8%;	Pred. No. 2;		
Matches 130;	Conservative 88;	Mismatches 203;	Indels 234;	Gaps 25;
QY	22	EKIKIESCSERTKIFRLYSVCTVEEKALILARONMRQKAVENLNRSLSLTHALLMWGASYL	81	
DB	960	QKLOLEKVITEAKKL-----EEQIILEDQNCCLAK	992	
QY	82	FDKLDHFHSSETPDSCGVFSQSIIMDGVIIHFSSILSGKGGNEEVKLCILLEAKHAQGY	141	
DB	993	-----EKKLEDRIAFTTNLT-----EEEKSKSLAKLKNKH	1025	
QY	142	SSDSTLFGEDHIKILDEESPNIFWSKLLGGKPNWKYPSPDTPORNKRQVYFEGSEASPK	201	
DB	1026	-----EAMITDLE-----ERLRREEKQKQDELE-----	1047	
QY	202	TGDBGNAKRRKASDDVTDPRVTDPPVDDDERKASGDHMGALESPKVITLOSSCKSSGT	261	
DB	1048	-----KTRRKLEGSDT-----LSDQIAELQAOIAELKMQWLAKKEE-----ELQALARVEE	1094	
QY	262	DGTLGDNDAGFLYSMGSHSIGIPEDMLASQDWCKIPDESORRLHTVLPKPMWKLQCVLHL	321	
DB	1095	EAAQKNWALKKRIELESQISELOQDESEASRNKAEKQKRDIGEELEAKLTELEPTLDS	1154	
QY	322	SDACTSMVGNFLYVJENHRIYEEPATTFQAFQI-----ALSWIAALL-----	364	
DB	1155	TAAQQLRSKREQEVNILKKTLEEEAKTHEA-QIQEMRQKHSQAVBELAQLEQTKRVKA	1213	
QY	365	-----VKQIL-----SHKESLV-----PANGELAF	384	
DB	1214	NLEKAKQTLNERGELANEVKVLLOGGDRSEHKRKKVQAQLQELQVKNFNEGERVVRTLAD	1273	
QY	385	KCSR--VEVDYIYIISLCSMKSLF-----LEHTQGLQFDFGTNSKQSVVSTK	429	
DB	1274	KVTQKLOVELDNVTGLLSQSDSKSKLTKPFSALESQLODTQELLQEE-----ENRQKLSLSTK	1330	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 10:23:56 ; Search time 5377.33 Seconds
(without alignments)
4395.843 Million cell updates/sec

Title: SEQ3 D1219_COPY_650_1270

Perfect score: 3217
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Delop 6.0			7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/cpn2.1/USPTO.spool/ROBINSON929/runat.11072005.102152.29186/app.query.fasta.1.2325
-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LQOPCI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=ROBINSON929 @CCN 1 1 8678 @runat.11072005.102152.29186 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gssi:
9: gb_gssi2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	413	12.8	654	8	BH989471
2	327.5	10.2	775	9	CL818501
3	317.5	9.9	764	9	CL795819
4	297.5	9.2	798	9	CG167203
5	296.5	9.2	369	8	BH419713
6	227	7.1	525	4	BG833433
7	224.5	7.0	569	2	BE203168
8	220	6.8	482	4	BG833434
9	214	6.7	668	5	BP133890

10	213.5	6.6	246	9	CC794080
11	211.5	6.6	516	1	AU288479
12	207.5	6.5	486	1	AI947519
13	200	6.2	799	8	B21272
14	198	6.2	560	8	BH788688
15	192	6.0	669	6	CA215647
16	183	5.7	942	8	B08967
17	181	5.6	529	5	BQ873535
18	179	5.6	969	9	CG396389
19	178.5	5.5	519	1	AU302228
20	168	5.2	690	8	BH978547
21	143.5	4.5	928	8	B11993
22	142.5	4.4	2079	3	AK017709
23	142.5	4.4	3927	9	CL971593
24	141.5	4.4	2265	3	AK014630
25	140.5	4.4	341	6	CB084203
26	138.5	4.3	1338	7	CO635655
27	138	4.3	2176	3	AK016726
28	137.5	4.3	3738	9	CL972961
29	135	4.2	2019	9	CL958706
30	134.5	4.2	848	8	BZ998864
31	134	4.2	7050	9	AY420513
32	133	4.1	3887	3	AK081567
33	133	4.1	4137	9	AY420057
34	132.5	4.1	681	7	CO561259
35	132.5	4.1	3180	9	CL961241
36	132.5	4.1	3807	3	AK029994
37	132.5	4.1	5048	3	HSM805203
38	131.5	4.1	1105	7	CR754231
39	131	4.1	1512	9	CL959694
40	131	4.1	2867	3	AK028516
41	131	4.1	5042	3	HSM805132
42	131	4.1	6062	3	HSM802948
43	131	4.1	6795	9	CL969745
44	130	4.0	668	6	CB526089
45	130	4.0	759	7	CK596760

ALIGNMENTS

RESULT 1

BH989471

LOCUS

DEFINITION

oem85a08.b1 B.oleracea002 Brassica oleracea genomic, genomic survey

sequence.

ACCESSION

BH989471

VERSION

BH989471.1

KEYWORDS

GSS.

SOURCE

Brassica oleracea

ORGANISM

Brassica oleracea

REFERENCE

1 (bases 1 to 654)

AUTHORS

Delenaunt, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,

Nash, W., Rabinowicz, P.D. and Wilson, R.K.

TITLE

Whole genome shotgun reads from Brassica oleracea

JOURNAL

Unpublished (2002)

COMMENT

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Plate: oem85 row: a column: 08

Seq primer: -21UPPT forward

Class: shotgun

High quality sequence start: 20

High quality sequence stop: 551.

Location/Qualifiers

1..654

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone_lib="B.oleracea002"

654 bp DNA linear GSS 07-OCT-2002

oem85a08.b1 B.oleracea002 Brassica oleracea genomic, genomic survey

sequence.

ACCESSION

BH989471

VERSION

BH989471.1

KEYWORDS

GSS.

SOURCE

Brassica oleracea

ORGANISM

Brassica oleracea

REFERENCE

1 (bases 1 to 654)

AUTHORS

Delenaunt, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,

Nash, W., Rabinowicz, P.D. and Wilson, R.K.

TITLE

Whole genome shotgun reads from Brassica oleracea

JOURNAL

Unpublished (2002)

COMMENT

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Plate: oem85 row: a column: 08

Seq primer: -21UPPT forward

Class: shotgun

High quality sequence start: 20

High quality sequence stop: 551.

Location/Qualifiers

1..654

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone_lib="B.oleracea002"

/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassaica oleracea TO1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Alignment Scores:
Pred. No.: 1,32e-34 Length: 654
Score: 413.00 Matches: 90
Percent Similarity: 67.95% Conservative: 16
Best Local Similarity: 57.69% Mismatches: 28
Query Match: 12.84% Indels: 22
DB: 8 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x BH989471 (1-654)

```
QY 358 SerTPIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuValArg 377
Db 253 AGTTGGATTGCGCTCGTCTGTAAGCAACGCTGAACGCGAAGAACTTTGGTCCGC 312
QY 378 AlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIle 397
Db 313 GCAAAATCGGAATTAGGTTTCAGTTGCTCTGGAAGAAGAGGTAGATTATATATATCTCT 372
QY 398 LeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPhe 417
Db 373 TTGTGCTGCATGAAGAGTATATTCCTGGGGCGCACACGAAGTTTGCGAGTGGATTACTTG 432
QY 418 GlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGly 437
Db 433 GGTACTAATCTTAAGCAACA-----AAGTGATGAAGATCTCTCAGGG 477
QY 438 AlaThrValArgAspGluLysIleLeuThrLysSerMetArgAsnSerSerGluAspGlu 457
Db 478 GATACAGTCGCTCAGGAACAGAGTAAACACAAATCGATCGCT-----519
QY 458 GluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLys 477
Db 520 -----AGCTCAGTAACAAAGATGTCGAAGAAG 546
QY 478 ThrIleSerGlyIleLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHis 497
Db 547 ACTATTAGCGACATCAAAATAAATGCAGTAAGAACTGCAGGAGCTTGTAACAATCCAC 606
QY 498 GluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGln 513
Db 607 GAGGAAAAAAGGTGGATCTGATGATAGTAGGAGTGTGCAAGAAGCAG 654
```

RESULT 2

CL818501 775 bp DNA linear GSS 09-AUG-2004
LOCUS OR_CBA0036H03.r OR_CBA Oryza rufipogon genomic clone OR_CBA0036H03
DEFINITION 3', genomic survey sequence.

ACCESSION CL818501

VERSION CL818501.1 GI:51064111

KEYWORDS GSS.

SOURCE Oryza rufipogon

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.

1 (bases 1 to 775)

Kim,H., Yu,Y., Watsoteki,M., Yost,D., Stum,D., Rao,K., Luo,M.,
Jettly,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
Wing,R.

OMAP project

Unpublished (2004)

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259

Email: <http://genome.arizona.edu>

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0036 row: H column: 03

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

FEATURES

source

1..775
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBA0036H03"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBA"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
drk treated 36 hrs before harvest"

ORIGIN

Alignment Scores:
Pred. No.: 4.9e-25 Length: 775
Score: 327.50 Matches: 85
Percent Similarity: 51.19% Conservative: 44
Best Local Similarity: 33.73% Mismatches: 104
Query Match: 10.18% Indels: 19
DB: 9 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x CL818501 (1-775)

```
QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 6 GATGCCATCATCATATATGGTTCTGCACCAACCCACTAAACGACTTGAAGGCTCTTCAA 65
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 66 AAGATCAGATAGAGTCACATTTGAGCGGTGTGACCATTTTTCGTTGTACACTCCTTTT 125
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
Db 126 ACAGTTGAGGAGAGAGTCTTGTTCTTCGAGGCGAGGCATTTGTTATTGATAACAACATA 185
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyValSerTyrLeuPhe 82
Db 186 CAAGATCTAAGGACTAGTTTAAACATTCCTTACTCAGATGGGGTGTGCTCATTTCTTTC 245
QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
Db 246 AGTAGCTGTAGTGGTCCAGCAAGATGACCATGCAAGTAAAGTTCTGAAATGAGAGA 305
QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlu 122
Db 306 CACTTTATTGATGAAGTAATCGTTGAGTTCTTTACAAAGTTTATCCACAACCTGTTGAAGAC 365
QY 123 GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
Db 366 AGCATGTAAGTACACAGGAAGTCCATAAGCAAGCAATATATGATGGGGAACTATATTCA 425
QY 143 SerAspSerThrLeuPheGlyLysAspHisIleLysLeuSerAspGluGluSerProAsn 162
Db 426 AGAATATATTACTCTGATGGGTGAAAGGAGGAATCTCTGTGCTTGGAGGATAATCCAGCT 485
QY 163 IlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182
Db 486 GAGTTCTGTTAAATTTATTGGATGGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 545
QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202
Db 546 CTGCANTCAAGATTACAAGTCCAGACCATGGATGAGGTCAATGTCTCCGGCTGAAGAA 605
```

QY 203 GlyAspGlyClyAsnAlaLysArgLys-----LysAlaSerAsp 216
 Db 606 ATCAATGAAGCCAGAAAGCGTAGAAAGGTTGGTGAAGATTATGGGTTTCATCTTCAAAA 665
 QY 217 AspValThrAspProArgValThrAsp----- 225
 Db 666 GTTGATCTACAAAAGTAGCAGTAGATGCGCTGATATTTGTACTACATCTGGCCCT 725
 QY 226 -----ProProValAspAspGluArgLysAla 235
 Db 726 GCTCTTCAACCACTGTGATGTCACGCAAAAATCA 761

RESULT 3

CL795819 764 bp DNA linear GSS 06-AUG-2004
 OR CBA0005M04.r OR CBA Oryza rufipogon genomic clone OR_CBA0005M04
 3', genomic survey sequence.

ACCESSION

CL795819

VERSION

CL795819.1 GI:51017835

KEYWORDS

GSS.

SOURCE

Oryza rufipogon

ORGANISM

Oryza rufipogon
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 764)
 Kim H., Yu Y., Wissotski M., Yost D., Stum D., Rao K., Luo M.,
 Jetty R., Kudrna D., Muller C., Hatfield J., Soderlund C. and
 Wing R.

TITLE

OMAP project

JOURNAL

Unpublished (2004)

COMMENT

Contact: Rod A. Wing
 Arizona Genomics Institute
 University of Arizona
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA
 Tel: 520 626 9595
 Fax: 520 621 1259
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: TAA TAC GAC TCA CTA TAG GG
 BACKWARD: CAC TCA TTA GGC ACC CCA
 Plate: 0005 row: M column: 04
 Seq primer: CAC TCA TTA GGC ACC CCA
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..764
 /organism="Oryza rufipogon"
 /mol_type="genomic DNA"
 /db_xref="taxon:4529"
 /clone="OR_CBA0005M04"
 /tissue_type="young leaves"
 /dev_stage="2 week old seedlings"
 /lab_host="DH10B T1 phage resistant"
 /clone_lib="OR_CBA"
 /notes="Vector: pAG1BAC1; Site 1: HindIII; Site 2: HindIII;
 drk treated 36 hrs before harvest"

ORIGIN

Alignment Scores:
 Pred. No.: 6,1e-24 Length: 764
 Score: 317.50 Matches: 84
 Percent Similarity: 50.00% Conservative: 42
 Best Local Similarity: 33.33% Mismatches: 107
 Query Match: 9.87% Indels: 19
 DB: 9 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x CL795819 (1-764)

QY 3 AspAlaPheLeuPheGlySerLeuAsnProSerHisAspValLysHisValGlu 22
 Db 6 GATGCCATCATATATATGTTCTGACAAACCCCACTAAACGACTTGAAGGCTTTCAA 65

QY 23 LysIleValIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
 Db 66 AAGATCAAGATAGAGTCACAAATTTGAGCGTGTGAGCAATTTTTCGCTGTACACTCCTTTT 125
 QY 43 ThrValGluGluLysAlaLeuIleLeuAlaAAGlnAsnMetArgGlnAsnLysAlaVal 62
 Db 126 ACAGTTTGAGGAGAGAGTCTTGTCTTGGAGCAGGCGCATTTGTTATTGATACACATTA 185
 QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
 Db 186 CAAATCTAAGGACTAGTTTAAACATCTCTACTCAGATGGGCTGTGGATGTGTTC 245
 QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerClyValSerPheGluGln 102
 Db 246 AGTAGACTGTGATGAGTCCAGCAAGTCCCAAGTAAAGTTCTGAAATGGAGAGA 305
 QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlu 122
 Db 306 CACTTTATTGATGAAGTAACTGTTGAGTCTTAAAGTTATCCAACTGTTGAAGAC 365
 QY 123 GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
 Db 366 AGCACTGAAGTACACAGGAAGTCCATAAGCAAGCTAATATGATGGGAACTATATTCA 425
 QY 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
 Db 426 AGAATATTACTCTCTGATGGTGAAGAGGAGGAATCTCTGTGTGAGGATAATCCAGCT 485
 QY 163 IlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182
 Db 486 GAGTCTCTGTTAAATTTATGATGGAAGATCCCTCATGCTGAGCTGTATATCTGAGCCA 545
 QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluLysProLysThr 202
 Db 546 CTGCAATCAAGATTACAAAGTCCAGACCATGATGAGTCAATGCTCGGCTGAAGAA 605
 QY 203 GlyAspGlyGlyAsnAlaLysLysArgLys----- 212
 Db 606 ATCAATGAAGCCAGCAAGAGCGTAGAAAGTTGGTGAAGTTATGAGTTTATGAGTTTCAAAA 665
 QY 213 -----LysAlaSerAspValThrAspProArgValThrAsp----- 225
 Db 666 GTTGATCTGACAAAGTAAACGATGATGATGCTGCTGATATTTGTACTACATCTGGCCT 725
 QY 226 -----ProProValAspAspGluArgLysAla 235
 Db 726 GCTCTTCAACCACTGTGATGTCACGCAAAAATCA 761

RESULT 4

CG167203

LOCUS

DEFINITION

genomic survey sequence.

ACCESSION

CG167203

VERSION

CG167203.1 GI:34058004

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

REFERENCE

1 (bases 1 to 798)

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

TITLE

Maize Genomics Consortium

JOURNAL

Unpublished (2003)

COMMENT

Other GSSs: PUIHJ72TB
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208

CG167203 798 bp DNA linear GSS 21-AUG-2003
 PUIHJ72TD ZM 0.6.1.0 KB Zea mays genomic clone ZMBBTA0583L24,
 genomic survey sequence.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

Maize Genomics Consortium
 Unpublished (2003)

Other GSSs: PUIHJ72TB
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

FEATURES

source
1..798
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0581L24"
/clone_lib="ZM_0.6_1.0 kb"
/note="Vector: PCR-toFO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

ORIGIN

Alignment Scores:
Pred. No.: 1,068-21 Length: 798
Score: 297.50 Matches: 74
Percent Similarity: 56.35% Conservative: 37
Best Local Similarity: 37.56% Mismatches: 77
Query Match: 9.25% Indels: 9
DB: 9 Gaps: 4

SEQ3_D1219_COPY_650_1270 (1-621) x CG167203 (1-798)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
DB 211 GATGCCATCATATATTGTTAGTACTGATCCATCCATGATGACCTTGAGCGCCCTTCAG 270
QY 23 LysIleLeuIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
DB 271 AGGGTCAGCATGAGTCACAGCTCGAACCTGTGCCTATTATTTTCGTTTGTACTCGTCTTT 330
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
DB 331 ACAGTGGGAGAAAGCTCTTATCTTGCAAGCATGACCATATTTTGGATAGTATATAT 390
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
DB 391 GTGATATAACCCCTAGCTGAGCCATTGCTGCTCAGTTGGGGCGCATCATTTCTCTTC 450
QY 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
DB 451 AATAGACTTTGAGGAGTTGCGCAACATAGTACTAC-----TCAAATGCTCTGCT 498
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
DB 499 GATGAGCTTTTCATGATTAATGTAGATTGGAGTTTTCACAAATTTACTCAGCAAGTT 558
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuAlaLysHisAlaGlnGlyThr 140
DB 559 GAATCAGAACTGAATCAGGCAATACAGCCATATCACAAGCTTATCTGTGATCTTTC 618
QY 141 TyrSerSerAspSerThrLeuPheGlyGlu----AspHisIleLysLeuSerAspGluGlu 159
DB 619 TATTTCTAGGCGCTATTCTTGTAGCGGTGAGAGGGAGGAATTCATCAGTTGATGTCAG 678
QY 160 SerProAsnIle-----PheTrpSerLysLeuGlyGlyLysAsnProMetTrpLys 177
DB 679 CTGCCAAAGTTCTGGGCCCTCTGTGTTAAGTTTACTTAATGGGAGGCTCTCTCAGTGGCAG 738
QY 178 TyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
DB 739 TATATATCTGAGCCAGTACAAAGAACCGCTAGAAAGATAAATAACATGAA 789

RESULT 5

BH419713

LOCUS

DEFINITION BH419713 369 bp DNA linear GSS 12-DEC-2001
BOGOE69TF BOGO Brassica oleracea genomic clone BOGOE69, genomic
survey sequence.

ACCESSION

BH419713

VERSION

BH419713.1

KEYWORDS

GI:17605441

SOURCE

ORGANISM

Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 369)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGOE69TR
Contact: Chris Town

TIGR Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

FEATURES

source

1..369
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGOE69"
/clone_lib="BOGO"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Alignment Scores:
Pred. No.: 3,898-22 Length: 369
Score: 296.50 Matches: 64
Percent Similarity: 72.36% Conservative: 25
Best Local Similarity: 52.03% Mismatches: 31
Query Match: 9.22% Indels: 3
DB: 8 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x BH419713 (1-369)

QY 483 LysLysLysTyrLysLysGlnValGlnLysLeuValGlnLysLysMet 502
DB 1 CAAAAAATGCAGTAAGCTCTGCAAGCTGTACAAACCTCGAGAGAAAGATG 60
QY 503 GluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerValGlu 522
DB 61 GACCTGATGATAGGAATGCTGTCAAGAGCAGGAACCTCAGAATTGTAAGGTGAA 120
QY 523 AlaAlaValIleArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeu 540
DB 121 GCATCATTTATTCGTGTACCTATTTCAGGTATATAATATCTCAGAGCTTACATGATGCTCTC 180
QY 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
DB 181 CAACGCGTGAATGATCTTTTCAAGAAAGATTGTATGATCTCTCAAGAGAGATTGGATGAA 240
QY 561 CysLeuLysSerLeuGlnMetHisAspValAlaLysLysLysLeuAlaGluAspGlu 580
DB 241 TGCCTTGAAGTTTAGACCAATAAACGAGGCTGGAAAGAGAGATTGGCTCAGAGATGAA 300
QY 581 AlaCysTrpIleAsnArgIleLysSerTrpAla----AlaLysLeuLysValCysValPro 599
DB 301 GCCTGTGGATTAGTCGGATAGAAATGGGCACGAGCTGAATTAAAGAAATGGTCTCTCC 360
QY 600 IleGlnSer 602
DB 361 AATCAAGCA 369

RESULT 6
BG833433
LOCUS
DEFINITION BG833433 525 bp mRNA linear EST 23-MAY-2001
951008C07.x2 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,

QY 226 ProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGlu 245
 Db |||||
 76 ---AATGTGGATCAGCCTTCATCAAAATCTGTAGTGAAGTACCCCAATGAATCAAG 132
 QY 246 SerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeu 265
 Db |||||
 133 GCAGAT-----AAACCTCAGGGTAAATGATGATGTTGAA 165
 QY 266 AspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGlu 285
 Db |||||
 166 TCCGGGAAAAC----- 177
 QY 286 AspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSerGlnArgArgLeuHis 305
 Db |||||
 178 -----GGCAGATCCATGATGAG---CTGAGGAGCTTACAT 210
 QY 306 ThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAlaCys 325
 Db |||||
 211 CTTTCGCTGAAGCCACAGATCAGAGCTCTGTGAAGTTCTTCTTCCAGATAATGTT 270
 QY 326 ThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu 345
 Db |||||
 271 AGAATGTGTTGCACAAATTTCTTGAGTATGTTACGAACAGCATCATTTCAATTGGGA 330
 QY 346 ProAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuVal 365
 Db |||||
 331 ---TCTGTTTCAATACAGTCTTTTCAAAATCTCTGATCAGGGCTGCTGCTTTGCTA 387
 QY 366 LysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLys 385
 Db |||||
 388 AATCAAAAGTTGACAGAGACTATCTTGCTCTTTTCAAAAGAACAAATGAAATTTTGAC 447
 QY 386 CysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPhe 405
 Db |||||
 448 TGTAGCAAGCAGAGAGTGACTATAGTATCTTTCTGAGTGTGAGAAATCTTT 507
 QY 406 LeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerVal 425
 Db |||||
 508 GTATATCGTACA-----GGCATTATATGATAATTTCTTCTCGAAAGTT 552

RESULT 8
 BG833434 482 bp mRNA linear EST 23-MAY-2001
 LOCUS 951008C07.x3 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,
 DEFINITION mRNA sequence.
 ACCESSION BG833434
 VERSION BG833434.1 GI:14191204
 KEYWORDS EST.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 482)

REFERENCE
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
 TITLE Walbot.V.
 JOURNAL University
 COMMENT Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 951008 row: C column: 07.
 Location/Qualifiers
 1. .482
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultiivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:4577"

FEATURES
 source

/tissue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth
 phases"
 /lab_host="DH10B"
 /clone_lib="951 - BMS tissue from Walbot Lab (GR)"
 /note="Vector: pUC19; Site.1: EcoRI; Site.2: EcoRI; The
 library was prepared by George Rudenko using poly (A)
 selected RNA and Universal Riboclone cDNA Synthesis System
 (Promega). cDNA was synthesized using both random and
 oligo(dT) primers in separate reactions and equipped with
 EcoRI adaptors. Library was size-fractionated on agarose
 gels (for insert size >400bp) and non-directionally cloned
 into EcoRI-digested pUC19 vector. Blue/white selection on
 carbenicillin-containing plates was used to recover
 positive clones."

ORIGIN

Alignment Scores:
 Pred. No.: 1.72e-13 Length: 482
 Score: 220.00 Matches: 50
 Percent Similarity: 64.29% Conservative: 22
 Best Local Similarity: 44.64% Mismatches: 34
 Query Match: 6.84% Indels: 6
 DB: 4 Gaps: 2

SEQ3_D1219_COPY_650_1270 (1-621) x BG833434 (1-482)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
 Db |||||
 142 GATGCCATCATCATATTTAGTAGTACATCCATGATGACTTGAGAGCCCTTCAG 201
 QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
 Db |||||
 202 AGGGTCAGCATGGAGTCACAGTCTGAACCTGTGCTCTATTTTCGTTTGTACTCGTCTTT 261
 QY 43 ThrValGluGluValAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaVal 62
 Db |||||
 262 ACAGTGGAGAAAGCTCTTATCTTGCAAGCATGACCATATTTTGGATAGTAAATTT 321
 QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
 Db |||||
 322 GTGAATATAACCCCTAGCTGAGCCATTGCTGCTCAGTTGGCGCATCATTTCTCTTC 381
 QY 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
 Db |||||
 382 AATAGACTTGGAGAGTTGCAGCAACATAGCTAC-----TCAAATGCTCTGCT 429
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
 Db |||||
 430 GATGAGCTTTTCATGATAATGATGATGAGTTT 465

RESULT 9

BG133890 668 bp mRNA linear EST 16-JUL-2003
 LOCUS BPI33890 MAT001 Nicotiana tabacum cDNA clone BV6204, mRNA sequence.
 DEFINITION BPI33890
 ACCESSION BPI33890
 VERSION BPI33890.1 GI:32876775
 KEYWORDS EST.
 SOURCE Nicotiana tabacum (common tobacco)
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiids; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 668)
 REFERENCE Matsuo, K., Tashiro, G., Horiguchi, T., Demura, T. and Fukuda, H.
 AUTHORS Profiling growth-phase dependent gene expression of tobacco BY-2
 TITLE cells by comprehensive microarray analysis
 JOURNAL Unpublished (2003)
 COMMENT Contact: Ken Matsuo
 Morphogenesis Research Group
 RIKEN Plant Science Center
 1-7-2 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9575

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT Contact: Taku Demura
Morphogenesis Research Group
RIKEN Plant Science Center
1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9605
Fax: 81-45-503-9573
Email: demura@postman.riken.go.jp
This clone was obtained at our laboratory.
Seq primer: M13 forward

FEATURES

source
1. .516
/organism="Zinnia elegans"
/mol_type="mRNA"
/cultivar="Canary bird"
/db_xref="taxon:34245"
/clone="Z2904"
/tissue_type="mesophyll cell"
/clone_lib="zinnia cultured mesophyll cell equalized cDNA"
/note="Vector: pGEM-T easy; cultured in tracheary element differentiation-inductive medium"

ORIGIN

Alignment Scores:
Pred. No.: 1.67e-12 Length: 516
Score: 211.50 Matches: 51
Percent Similarity: 58.33% Conservative: 19
Best Local Similarity: 42.50% Mismatches: 46
Query Match: 6.57% Indels: 4
DB: 1 Gaps: 1

SEQ3_D1219_COPY_650_1270 (1-621) x AU288479 (1-516)

QY 299 GluSerGlnArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnVal 318
DB 82 GAAGGGAAATGAACCTTTAAGACCTCTGAAGCTTAAATATACCTAAACTATGTGAAGTC 141
QY 319 LeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
DB 142 CTTAAGTTTCGGAGGATGTGAAGATAATGGTGAAGGATTTCTCGAGTATGTTATTGAG 201
QY 339 AsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleLeuSer 358
DB 202 AATTATCGGTCAACGAAGCACTGAAGACACATTACATGCTTCTTGATATCCCTGTGT 261
QY 359 -TrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAl 378
DB 262 TNGGATTGTTCCGGATTGTAAACACAAAGCTTGATAGGAAGCAATCCATAGTCTTGC 321
QY 378 aAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLe 398
DB 322 GAACGAACACTTGGGTTTCAGTTGTGAAGGAAATGAAGCAACTCAGTCTACTCGAAGCT 381
QY 398 uSerCysMetLysSerLeuPheLeuGluHisThr-----GlnGlyLeuGlnPhe 414
DB 382 GGAATTTGCAAAAGAGATGTTCTTACTTCACAGGACATCCCAAGAAACCTTCGTTTC 439

RESULT 12

LOCUS A1947519 486 bp mRNA linear EST 19-AUG-1999
DEFINITION 603022G06.x1 603 - stressed root cDNA library from Wang/Bohnert lab
Zea mays cDNA, mRNA sequence.

ACCESSION A1947519.1 GI:5739829
VERSION EST.
KEYWORDS Zea mays
SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 486)

REFERENCE

AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL COMMENT

University Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603022 row: G column: 06.

FEATURES

source
1. .486
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/clone_lib="603 - stressed root cDNA library from Wang/Bohnert lab"
/note="Organ: root; Vector: pBluescriptII SK(+) XR;
Seedling stressed root cDNA library from Wang/Bohnert lab"

ORIGIN

Alignment Scores:
Pred. No.: 4.19e-12 Length: 486
Score: 207.50 Matches: 56
Percent Similarity: 53.21% Conservative: 27
Best Local Similarity: 35.90% Mismatches: 64
Query Match: 6.45% Indels: 9
DB: 1 Gaps: 4

SEQ3_D1219_COPY_650_1270 (1-621) x A1947519 (1-486)

QY 44 ValGluGluLysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnLysAlaValGlu 63
DB 1 GTGGAGGAAAAGCTCTTATTTCTGCAAGCATGACCATATTTTGGATAGTAAATTTGTG 60
QY 64 AsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAsp 83
DB 61 AATAAACCCTTAGCTGAGCCATTGTCTGCTCAGTTGGGGGCGCATCATTTCTTCTCAT 120
QY 84 LysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPheGlu 101
DB 121 ACACCTTGAGGAGTTGCAGCAACATAGTAC-----TCAAATCTCTCTGTGTAT 168
QY 102 GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGly 121
DB 169 GAGCTTTTCATGATAATGTAGATTGGAGTTTTCACAAATATCTCAGCAAGGTGAA 228
QY 122 GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141
DB 229 CTCGAACCTGATCAGGCAATACACCATATCACAAGCTTATCTGTGTGATCTTCTAT 288
QY 142 SerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGluGluSer 160
DB 289 TCTAGGGCTATTGTTGTAGCGGTCGAGAGGAGGAAATTCATCAGTTGATGTCACCTG 348
QY 161 ProAsnIle-----PheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyr 178
DB 349 CCAAAAGTTCTGGCGCTACTGGTTAAGTTTACTTAATGGGAGGTCTCCTCAGTGGCAGTAT 408
QY 179 ProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
DB 409 ATATCTGAGCCAGTACAAAGAGCGGTAGAAAGATAAATAACATGAA 456

RESULT 13

B21272 799 bp DNA linear GSS 16-SEP-1997
LOCUS F3G11-T7.1 IGF Arabidopsis thaliana genomic clone F3G11, genomic
DEFINITION survey sequence.
ACCESSION B21272

Db	139	TCAAATGACTTGGAGAGTTGCAGCAACATAGTAC-----TCAAATGTCTCTG	92	491	AGCTCAAGATAGGTCACAACTAGATGTGGCATTTTCGGCTTATATACTCTCTTC	555
Qy	100	heGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLys	119	Qy	43	ThrValcIuGluIysAlaLeuIleLeuAlaArgGlnAsnMetArgGlnAsnIysAlaVal
Db	91	GTGATGAGCTTTTCATGGATAATGTAGATTTGGAGTTTTTGACAAAAATTTACTCAGCAAG	33	Db	551	ACATGTGAGAGAGAGAGATCTTGTGCTCGCCAAAGACGATGATTTATTGATAGCAACAGC
LOCUS	CA215647	669 bp	EST 25-SEP-2003	Qy	63	GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu
LOCUS	CA215647	669 bp	EST 25-SEP-2003	Db	611	CAAGACATACATCTAGTTTAAAGCCATTGCTTGGTCAGCTGGGGGTATCATTCCTC

```
Search completed: July 11, 2005, 23:52:32
Job time : 5382.33 secs
```

```

ORIGIN
  Occurred at http://success.rax.rax.uniCamp.52/pwz.rax
  Alignment Scores:
  Pred. No.: 3.64e-10 Length: 669
  Score: 192.00 Matches: 37
  Percent Similarity: 73.42% Conservative: 21
  Best Local Similarity: 46.84% Mismatches: 21
  Query Match: 5.97% Indels: 0
  DB: 6 Gaps: 0
  SEQ3_D1219_COPY_650_1270 (1-621) x CA215647 (1-669)
  Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyHisValGlu 22
  Db 431 GATTCCATTATCATTTATGTAGTACTTGAATCCACTAAATGATCTGAAGGCTTTCGA 490
  Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
  Db 431 GATTCCATTATCATTTATGTAGTACTTGAATCCACTAAATGATCTGAAGGCTTTCGA 490

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 19:44:24 ; Search time 61.6667 Seconds
(without alignments)
5156.776 Million cell updates/sec

Title: SEQ3_D1219_COPY_650_1270
Perfect score: 3217
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3213	99.9	2001	2 Q9M659	Q9m659 arabidopsis
2	3211	99.8	2001	2 Q9M658	Q9m658 arabidopsis
3	3211	99.8	2254	2 Q9LN02	Q9ln02 arabidopsis
4	163.5	5.1	1461	2 Q17909	Q17909 caenorhabdi
5	160.5	5.0	842	2 Q6ZU43	Q6zu43 homo sapien
6	160.5	5.0	1738	2 Q9SI41	Q9si41 arabidopsis
7	160.5	5.0	1739	1 CHD2_HUMAN	Chd2 human
8	158.5	4.9	1268	2 Q7RG05	Q7rg05 plasmodium
9	158	4.9	1245	2 Q8IAN1	Q8ian1 plasmodium
10	157	4.9	1320	2 Q9JK25	Q9jk25 rattus norv
11	157	4.9	4007	2 Q7SHZ4	Q7shz4 neurospora
12	156.5	4.9	1330	2 Q22342	Q22342 caenorhabdi
13	155	4.8	905	2 Q6BHF8	Q6bnf8 debaryomyce
14	154.5	4.8	1033	2 Q676A5	Q676a5 oikopleura
15	153.5	4.8	2230	1 GOA4_HUMAN	G0a4 human
16	153	4.8	1611	2 Q7RD43	Q7rd43 plasmodium
17	151	4.7	2757	2 Q7RRR9	Q7rrr9 plasmodium
18	150	4.7	979	2 Q9EP71	Q9ep71 mus musculus
19	150	4.7	1786	2 Q7ZT34	Q7zt34 poephila gu
20	149.5	4.6	1310	2 Q949K0	Q949k0 lycopersico
21	149	4.6	2067	2 Q9U3U8	Q9u3u8 plasmodium
22	149	4.6	2160	2 Q8IE50	Q8ie50 plasmodium
23	148.5	4.6	2042	2 Q6BUQ9	Q6buq9 debaryomyce
24	148.5	4.6	7210	2 Q9V7G8	Q9v7g8 drosophila
25	148.5	4.6	9270	2 Q8MJD9	Q8mid9 drosophila
26	148	4.6	1500	2 Q75J40	Q75j40 cryza sativ
27	147.5	4.6	992	2 Q9NTH6	Q9nth6 homo sapien
28	147.5	4.6	1927	2 Q75WX9	Q75wx9 helicobacte
29	147.5	4.6	1938	2 Q9PVE1	Q9pve1 brachydanio
30	146	4.5	1456	2 Q9UPV0	Q9upv0 homo sapien
31	146	4.5	1939	1 MYH4_HUMAN	Myh4 human

RESULT 1

ID	Q9M659	PRELIMINARY;	PRT;	2001	AA.
AC	Q9M659;				
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	MOM.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-20279299; PubMed-10821279; DOI=10.1038/35012108;				
RA	Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;				
RT	"Disruption of the plant gene MOM releases transcriptional silencing				
RT	of methylated genes.";				
RL	Nature 405:203-206(2000).				
DR	EMBL; AF213627; AAF73380.1; --				
SQ	SEQUENCE 2001 AA; 218585 MW; F6C5DAE897F2FE61 CRC64;				
Query Match 99.9%; Score 3213; DB 2; Length 2001;					
Best Local Similarity 99.8%; Pred. No. 5.1e-171;					
Matches 620; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
QY	1	RADAFILFGSSLNPSHDVXHKIESCSERTKIFRLYSVCTVEEKALILARQNRRNK	60		
Db	650	RADAFILFGSSLNPSHDVXHKIESCSERTKIFRLYSVCTVEEKALILARQNRRNK	709		
QY	61	AVENLNRSILTHALLMWGASYLFDKLDHFHSSSTPDSGVSEQSIMDGVIFEFSILSSKG	120		
Db	710	AVENLNRSILTHALLMWGASYLFDKLDHFHSSSTPDSGVSEQSIMDGVIFEFSILSSKG	769		
QY	121	GEENEVKLCLEAKHAQTYSSDSTLFGEDHIKLSDESPNIFWSKLLGGKNPMWKYPS	180		
Db	770	GEENEVKLCLEAKHAQTYSSDSTLFGEDHIKLSDESPNIFWSKLLGGKNPMWKYPS	829		
QY	181	DTPORRKRVQVFESEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH	240		
Db	830	DTPORRKRVQVFESEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH	889		
QY	241	MGALESFKVITLQSSCKSGTGTDLGNDAGFLYSMGSHISGIPEDMLASQDMGKIPDES	300		
Db	890	MGALESFKVITLQSSCKSGTGTDLGNDAGFLYSMGSHISGIPEDMLASQDMGKIPDES	949		
QY	301	QRRHLTVLKPMAKLCQVHLSDACTSMVGNFLEYVNIENRIYEEPATTFQAFQIALSWI	360		
Db	950	QRRHLTVLKPMAKLCQVHLSDACTSMVGNFLEYVNIENRIYEEPATTFQAFQIALSWI	1009		
QY	361	AALLVKQILSHKESLVANSELAFCRSRVVDYIYILSCMSLFLFHTQGLQFCFGTN	420		
Db	1010	AALLVKQILSHKESLVANSELAFCRSRVVDYIYILSCMSLFLFHTQGLQFCFGTN	1069		

ALIGNMENTS

QY 421 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 480
 DB 1070 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 1129
 QY 481 GIKKKYKQVOKLVQEHHEKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 540
 DB 1130 GIKKKYKQVOKLVQEHHEKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 1189
 QY 541 KLLDHNRYERKFDIEIKSEKNECLSKLEOMHDVAKKLADEACWINRIKSWAAKLVKCVPI 600
 DB 1190 KLLDHNRYERKFDIEIKSEKNECLSKLEOMHDVAKKLADEACWINRIKSWAAKLVKCVPI 1249
 QY 601 QSGNNKHFSGSSNISQNPV 621
 DB 1250 QSGNNKHFSGSSNISQNPV 1270

RESULT 2

Q9M658 PRELIMINARY; PRT; 2001 AA.
 AC Q9M658
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE MOM.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=2079299; PubMed=10821279; DOI=10.1038/35012108;
 RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
 RT "Disruption of the plant gene MOM releases transcriptional silencing
 RT of methylated genes";
 RL Nature 405:203-206(2000).
 DR EMBL; AF213628; AAF73381.1; --
 SQ SEQUENCE 2001 AA; 218567 MW; B9C85D4B6270441 CRC64;

Query Match 99.8%; Score 3211; DB 2; Length 2001;
 Best Local Similarity 99.8%; Pred. No. 6.7e-171;
 Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RADAFILFGSSLNPSHDVVKHVEKIKIESCSERTKIFRLYSVCTVEKALILARQNQRN 60
 DB 650 RADAFILFGSSLNPSHDVVKHVEKIKIESCSERTKIFRLYSVCTVEKALILARQNQRN 709
 QY 61 AVENLNRSITALLMWGASYLFDKLDHFSSTPDSGVSEFQSIMDGVHEFSSILSSKG 120
 DB 710 AVENLNRSITALLMWGASYLFDKLDHFSSTPDSGVSEFQSIMDGVHEFSSILSSKG 769
 QY 121 GENEVKLCILLEAKHAQTSYSDSTLFGEDHILKLSDESPNIFWSKLLGGKNPMWKYPS 180
 DB 770 GENEVKLCILLEAKHAQTSYSDSTLFGEDHILKLSDESPNIFWSKLLGGKNPMWKYPS 829
 QY 181 DTPQRNRKVOYFEGSEASPDKGDNNAKRRKASDDVTDPRVTDPPVDDRRKASGKDH 240
 DB 830 DTPQRNRKVOYFEGSEASPDKGDNNAKRRKASDDVTDPRVTDPPVDDRRKASGKDH 889
 QY 241 MGALESKPVITLQSSCKSGTDTLGDNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 300
 DB 890 MGALESKPVITLQSSCKSGTDTLGDNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 949
 QY 301 QRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVIEHRIYEEPAITTFQAFIALSWI 360
 DB 950 QRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVIEHRIYEEPAITTFQAFIALSWI 1009
 QY 361 AALLVKQILSHKESVRANSELAFCRSRVEVDYIYSILSCMKSLFLEHTQGLQFCFGTN 420
 DB 1010 AALLVKQILSHKESVRANSELAFCRSRVEVDYIYSILSCMKSLFLEHTQGLQFCFGTN 1069

QY 421 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 480
 DB 1070 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 1129
 QY 481 GIKKKYKQVOKLVQEHHEKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 540
 DB 1130 GIKKKYKQVOKLVQEHHEKKMELNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 1189
 QY 541 KLLDHNRYERKFDIEIKSEKNECLSKLEOMHDVAKKLADEACWINRIKSWAAKLVKCVPI 600
 DB 1190 KLLDHNRYERKFDIEIKSEKNECLSKLEOMHDVAKKLADEACWINRIKSWAAKLVKCVPI 1249
 QY 601 QSGNNKHFSGSSNISQNPV 621
 DB 1250 QSGNNKHFSGSSNISQNPV 1270

RESULT 3

Q9LN02 PRELIMINARY; PRT; 2254 AA.
 AC Q9LN02
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE T6D22.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thavari A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thavari A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC026875; AAF79839.1; --
 DR PIR; D86215; D86215.
 DR HSP; P00918; 1BV3.
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
 DR InterPro; IPR01148; Euk COandh.
 DR InterPro; IPR011009; Kinase like.
 DR Pfam; PF00194; Carb anhydrolase; 1.
 DR ProDom; PD000865; Euk COandh; 1.
 SQ SEQUENCE 2254 AA; 248157 MW; 64E80CF274791PF3 CRC64;

Query Match 99.8%; Score 3211; DB 2; Length 2254;
 Best Local Similarity 99.8%; Pred. No. 7.8e-171;
 Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RADAFILFGSSLNPSHDVVKHVEKIKIESCSERTKIFRLYSVCTVEKALILARQNQRN 60
 DB 883 RADAFILFGSSLNPSHDVVKHVEKIKIESCSERTKIFRLYSVCTVEKALILARQNQRN 942
 QY 61 AVENLNRSITALLMWGASYLFDKLDHFSSTPDSGVSEFQSIMDGVHEFSSILSSKG 120
 DB 943 AVENLNRSITALLMWGASYLFDKLDHFSSTPDSGVSEFQSIMDGVHEFSSILSSKG 1002

QY 121 GEENEVLCILLLEAKAAGTYSSTDLFGEDHKLSDSESPNIFWKLGGKKNPMWKYPS 180
 DB 1003 GEENEVLCILLLEAKAAGTYSSTDLFGEDHKLSDSESPNIFWKLGGKKNPMWKYPS 1062
 QY 181 DTPQRNKRVOYEGEGEASPKTGGGNNAKRRKASDDVTPRVTDPVDDDERKASGKDH 240
 DB 1063 DTPQRNKRVOYEGEGEASPKTGGGNNAKRRKASDDVTPRVTDPVDDDERKASGKDH 1122
 QY 241 MGALESFKVITLQSSCKSGTDTLGDNDAGLYSMGSHISGIPEDMLASQDMGKIPDES 300
 DB 1123 MGALESFKVITLQSSCKSGTDTLGDNDAGLYSMGSHISGIPEDMLASQDMGKIPDES 1182
 QY 301 ORRLHVLTPKMAKLCQVHLSDACTSMVGNFLEYVIEHRIYEEPAITFOAQIALSWI 360
 DB 1183 ORRLHVLTPKMAKLCQVHLSDACTSMVGNFLEYVIEHRIYEEPAITFOAQIALSWI 1242
 QY 361 AALLVKQILSHKESLVRANSELAFKCSRVEVDVYIYILSCMKSLFLEHTQGLQDFCGTN 420
 DB 1243 AALLVKQILSHKESLVRANSELAFKCSRVEVDVYIYILSCMKSLFLEHTQGLQDFCGTN 1302
 QY 421 SKQSVVSTKLVBNSLGSATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIS 480
 DB 1303 SKQSVVSTKLVBNSLGSATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIS 1362
 QY 481 GIKKKYKQVQKLVQHEHEKQKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 540
 DB 1363 GIKKKYKQVQKLVQHEHEKQKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGD 1422
 QY 541 KLLDHNVERKFDIEKSEKNECLSLQMDHDAKCKLAEDACWINRIKSWAAKLKVCVPI 600
 DB 1423 KLLDHNVERKFDIEKSEKNECLSLQMDHDAKCKLAEDACWINRIKSWAAKLKVCVPI 1482
 QY 601 QSGNKHFGSSNISQNPV 621
 DB 1483 QSGNKHFGSSNISQNPV 1503

RESULT 4

O17909 PRELIMINARY; PRT; 1461 AA.
 AC O17909;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE Hypothetical protein H06001.2.
 GN ORFNames=H06001.2;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Barlow K.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: Contains 2 chromo domains.
 DR EMBL; Z292970; CAB07481.2; -.
 DR FIC; T23056; T23056.
 DR WormBase; WBGene00010369; H06001.2.
 DR WormPep; H06001.2; CE32454.
 DR GO; GO:0000785; C:chromatin; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0003682; F:chromatin binding; IEA.
 DR GO; GO:0004386; F:helicase activity; IEA.
 DR GO; GO:0006333; F:chromatin assembly or disassembly; IEA.
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR002464; DEAH box.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00385; Chromo; 2.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00176; SNF2_N; 1.
 DR SMART; SM00298; CHROMO; 2.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICC; 1.
 DR PROSITE; PS00598; CHROMO 1; 2.
 DR PROSITE; PS0013; CHROMO 2; 2.
 DR PROSITE; PS00690; DEAH ATP HELICASE; UNKNOWN_1.
 KW Hypothetical protein; Nuclear protein.
 SQ SEQUENCE 1461 AA; 168998 MW; 0DEA2BBSBAB91958 CRC64;
 Query Match 5.1%; Score 163.5; DB 2; Length 1461;
 Best Local Similarity 17.6%; Pred. No. 1.5;
 Matches 136; Conservative 95; Mismatches 236; Indels 257; Gaps 27;
 QY 2 ADAPILFGSLNPSHDVKEVKEIKIESCSERTKIFRLYVCTVVEKALILARQNMQKA 61
 DB 799 ADTVAFIDSDWNPQNDQMSRAHRIGQTKTVNIYRLVTKGSVEE-----EI 845
 QY 62 VENVLRSLTHALLMGASVLPDKLDH-----FHSETPDGSPVSPQSTMDGV 108
 DB 846 VERAARKLV-----LDHLVTQMDTGKTVLSKNATASGSPVDPKQLSAI 891
 QY 109 IHEFSSI--LSSKGGSENEV-----KLCILLLEAKAAGTYSSTDLFGSE-DHIKLSDES 160
 DB 892 L-KFGAVELFKEGEQEPEVDIDRLMGAEATREAEVEEMKENELSSPKYANFAIDER 950
 QY 161 PNIPIF-----WSKLGKGNPMWKYPSDTPQNRKV-----Q 191
 DB 951 KDIAAATDEWAAII-----PEEDNRILEERMKELAEMLNAPRQKQPIQ 997
 QY 192 YFEGSEA-SPKTGCGNNAKRRKASDDVTPRVT-----TDPVDDDERK 234
 DB 998 VVEDDDGDDDEEDDTGKKKKKAGVGNFTIPEIKRPIKSRKPSMPLNLEEIAQAEUE 1057
 QY 235 ASGKDHMGALSPKVVITLQSSCKSGTDTLGDNDAGLYSMGSHISGIPEDMLASQDMG 294
 DB 1058 EHSITDEMKG-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
 QY 295 KIPDSQ-----RRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYV 336
 DB 1093 AAESKQOIERKFKFHTCDVNLKQIERSHAELKP-----LHEILKSEETKTS 1139
 QY 337 IENHRIYEEPAIT-----FQAFOIAL-----SWIAALLVKQI-LSHKESL 375
 DB 1140 -----FKPPANAKLQKGDVDMWRPDDSAALLGVWKYGYGSWEAIKMDPTGLADKIFI 1193
 QY 376 VRANSELAFKCSRVEVDVYIYILSCMKSLFLEHTQGLQDFCGTNSKQSVSTKLVBNSL 435
 DB 1194 KDTKKKPGQKNLQVADVILKLM-----SKQVKYTE-----1225
 QY 436 SGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIKIKKKYKQVQKLVQ 495
 DB 1226 -----KKERKRKADDPVGVPEKKKRHTNNVPQEGE-----KKKEKK-----1262
 QY 496 EHEEKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFEDEK 555
 DB 1263 --BEKNSSSLK---DQLALLSIDKSLYXGALEDSAKPFLECVKLCMPVHRTMKLKEAQ 1317
 QY 556 SEKNELCKS-----LEOMHDVAKCKLAEDACWINRIKSWAAKLKVCV 599
 DB 1318 EAKNQADEAKYLTRLGDSFLENLETLIKKKPKTNIRKWNYLWIFLCKFTLREP 1371

RESULT 5

```
Q6ZU43
ID Q6ZU43 PRELIMINARY; PRT; 842 AA.
AC Q6ZU43
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein FLJ44007.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
OX [1]
RN
RP
SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saico K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagehara K.,
RA Masuho Y., Nagai K., Isegai T.;
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AKI25995; BAC86384.1; -.
SQ SEQUENCE 842 AA; 94360 MW; F195761EFD08E33B CRC64;

Query Match 5.0%; Score 160.5; DB 2; Length 842;
Best Local Similarity 20.1%; Pred. No. 1.1;
Matches 123; Conservative 98; Mismatches 221; Indels 169; Gaps 26;

QY 122 ENEVKLCILLBAKHAQGYTSDSTLFGEDHIKLSDESPNTFWSKLGG-----171
Db 245 EKSEPKICRNLTTPKADPT-----GSEPAKASEKAPE---DTVDAGEGSRREAA 293
QY 172 KNPWKY-----PSTPQ-----RNRKRVQVPEGSEAPKTDGNGNAKRRKASDDVTD 220
Db 294 KEPPKASALEGSSDASQELISEHMKPQLSDSIASDPKSPHGLDFGRGRISHLLD 353
QY 221 PRVTD-----PVDDERKASGDHMGAL-----PKVITLQ 253
Db 354 VDVLSPVLGACRQAQPIGIEDKDSQSDELQSKGLSEERYHRLSPPLPHEERAQ 413
QY 254 SCKS-----SGTDGTL-----GNDAPGLYSGMHSIG-----IPEDMLA--- 289
Db 414 SPRLSATEBEPQGPQEPQEWKEAELGSDSAALSLSLQSLQRRSTPVPAPPEQLSEAA 473
QY 290 ----SQDWKIPDESORRLHTVLKPM-----AKLQ-----VLHL-----SDACTSMYGNEL 333
Db 474 LKAMEEAVAVQLEQDQRHLLSKQEKMQQLREKLCQEEBEEILRLHQKEQSLSSURRL 533
QY 334 EYVIENHRIYEPATTFOAFQALISWIAALLVKQLLSHKSILVRANSSELAFAKCSRVEDY 393
Db 534 QKAIE-----EEARMEESQRLSWLA-QVQSSTQADEQIRAEQASLQKLRELE- 586
QY 394 IYSLSCMK-----SLFLEHTQGL-----QDFCGTNSKQSVST-----KLVNESLSGA 438
Db 587 ----SQKAERASLQKRMQLQKLEIEA-SEKSEQAALNAAKEKALQQLREQLGE 640
QY 439 TVRDEKINTKSMNSSEDECEWTEKCSHYSTATRDI-----EKTISGKKYKKVOKLV 494
Db 641 --RKEAVATLEKHSAB-----LERLCSSLEAKHREVSVSLQKIQEAOQKEAQKCL 693
QY 495 QEHEEK-----KMLLNMYADKKQKLE-----TSKSVEA 523
Db 694 GQVEHVRHOKSVHVAQVEHLSLLREKQVEGEHERLDRKMEHQVMAKAREQVEA 753
QY 524 AVIRITCSTSTQVGLKLLDHNRYKFDIEKSEKNECLKSLQEMQHDVAKKLAEDEACW 583
Db 754 EERKQRAELGLHTGLERLQRAHERELETVRQEHKRLFDLRRRREOERKLQDLLELD 813
QY 584 INRIKSWAAKL 594
Db 814 ETRAKDVKARL 824
```

RESULT 6

```
Q9S141
ID Q9S141 PRELIMINARY; PRT; 1738 AA.
AC Q9S141
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative chromodomain-helicase-DNA-binding protein.
NCBI_TaxID=3702;
GN Name=At2g13370;
OS Arabidopsis thaliana (Mouse-ear cress)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurooids II; Brassicales; Brassicaceae; Arabidopsi.
OX [1]
RN
SEQUENCE FROM N.A.
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buehl C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RN
SEQUENCE FROM N.A.
RA Town C.D., Kaul S.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Contains 2 chromo domains.
DR EMBL; AC007209; RAD28668.1; -.
DR PIR; C84507; C84507.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003682; F:chromatin binding; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006333; P:chromatin assembly or disassembly; IEA.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; Chromo; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS50013; CHROMO_2; 2.
DR DNA-binding; Helicase; Nuclear protein.
SQ SEQUENCE 1738 AA; 198695 MW; CE50069DD8C3B6CD CRC64;

Query Match 5.0%; Score 160.5; DB 2; Length 1738;
Best Local Similarity 18.5%; Pred. No. 2.7;
Matches 138; Conservative 104; Mismatches 234; Indels 269; Gaps 31;

QY 2 ADATILFGSSLNPSHDVKHVEKIKIESCSERTKIPRLYSVCTVEKALILARQM-----56
Db 1041 ADTVVIFSDWNPQDLQMSRAHRIGQEVVNIYRFVTSKSVSEBILERAQRKQVLDHL 1100
QY 57 -----RQKAVENLRSLTHALLMWGASYLPDKLDHFSHSETPDGVSFQS 103
Db 1101 VTQKLNABGRLEKRETKKSGNFDKNELSAIRLFGABELFKE-----1141
QY 104 IMDGVITHEFSSILSSKGGSENEVKLCLL-----LEAKHAQGYTSDSTLFGEDH 152
Db 1142 -----DKNDESKRLLSMDIDILERAQVEEKH---TDTEHELLGAFK 1184
QY 153 I-KLSDSESPNIFWKLGLGKNPMKYPSTDPQNRKRVQVYFEGSEASPTDGGNAKRR 211
Db 1185 VANFCAEDDGSFWSRWIKPDSVVTAAEALAPRAARNTKSYVDPSPHPD-----RTSKRK 1238
QY 212 KKASDDVTDPRVTDPPVDDDERKASGDHMGALSKPVITLQSSCKSGTGDGTLGDNAF 271
```

Db 1239 KKGSE-----PPEHTERSQKRKRTYFVPSTPLLEGTSAQVRGWSYGNLPKRDAQ 1288
QY 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHTVLKPKWAKLCQV 318
Db 1289 RFYRTVTKFNGHNMQACIABEVGVVE-----AAPEEAQVELFDA-----1328
QY 319 LHLSDACTSNV--GNF-----LEVVHENRIYEEPPATTFQ 351
Db 1329 --LIDGCKESVETGNFEPKGPVLDFGVPVKANELLKRVQGLQSLKRSRYNDPISQFR 1386
QY 352 AFQIALSWIA-----ALLVKQIILSH--KESLVANSELAF--KCSRVVEV 391
Db 1387 ----VLSYKPSNWSKCGNQNQDDARLLGLYHFGWKEKIRLDESGLTKKIAPVEL 1442
QY 392 DYIYSILSKMSKSLFLEHTQLODFCF-----GTNSKQSVVSTKLVNESLSG---ATVRDE 443
Db 1443 QHHTETPLRAPNLKERATALLAMELAAAGGKNTAKASRKNKSKVDNLINQFKAPARDR 1502
QY 444 K-----INTK-----SWRNSS-----DEE 458
Db 1503 RGKSGPANVSLSTKDGPRKTQKAEPLVKEEGEMSDGVEYQFKEQKWMWECEDVLADE 1562
QY 459 CMTEKCSHYSTATRDI--EKTISGIKKYK---KQVKLVQSHEE-----KMWELN 506
Db 1563 IKTLGRQLRLQTSADLPKEKVKFKIRYLEILGRIDAIVLSEHEDLYKQDRMTWRLWN 1622
QY 507 -----MYADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLLDHNTERKF 551
Db 1623 YVSTFNSLGDRLNQIYSLKQKEEVEGVGPS--HLNGSR-----NQFQ--1666
QY 552 DEIKSENECKLSLEQMH---DVAK 573
Db 1667 QKFPTAGNS--QGSQQVHKGIDTAK 1689

RESULT 7

CHD2_HUMAN STANDARD; PRT; 1739 AA.
ID CHD2_HUMAN
AC O14647;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein 2 (CHD-2).
GN Name=CHD2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97470991; PubMed=9326634; DOI=10.1073/pnas.94.21.11472;
RA Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT "Characterization of the CHD family of proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC -!- FUNCTION: Sequence-selective DNA-binding protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF006514; AAB87382.1; -.
CC HSSP; P23197; IAP0.
CC Genew; HGNC:1917; CHD2.
CC MIM; 602119; -.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0007001; P:chromosome organization and biogenesis (gen. . .; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR InterPro; IPR00953; Chromo.
DR InterPro; IPR002464; DEAH box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; Chromo; 2.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR PROSITE; PS00598; CHROMO 1; 2.
DR PROSITE; PS0013; CHROMO 2; 2.
DR PROSITE; PS00690; DEAH ATP HELICASE, FALSE NEG.
KW ATP-binding; DNA-binding; Helicase; Hydrolase; Nuclear protein;
Repeat.
KW DOMAIN 1 74 Ser-rich.
FT DOMAIN 121 139 Ser-rich.
FT DOMAIN 261 353 Chromo 1.
FT DOMAIN 378 456 Chromo 2.
FT NE_BIND 509 516 ATP (Potential).
FT SITE 617 620 DEAH box.
SQ SEQUENCE 1739 AA; 200560 MW; D62089C25EBA440F CRC64;
Query Match 5.0%; Score 160.5; DB 1; Length 1739;
Best Local Similarity 18.9%; Pred. No. 2.7;
Matches 152; Conservative 115; Mismatches 260; Indels 279; Gaps 35;
QY 2 ADAPILFGSLNPSHDVGVKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVVIFSDWNPQDLQAARAHRIQKKQVNIYRLVTGTVVEEIERAKKMWLDHL 938
QY 50 ILARQNRRQKAVEN-----LNRLTHALLMMGASVLPKLFHSHSETPDGVSF 100
Db 939 VIQRMDTTGRTTILENNSGRSNPNFKEELTAILKFGAEDLFKELE-----GEES 988
QY 101 EQSINDGVIEHFSSILSKGGENEVKLCILLEAKHAQGYSSDSTLFGEDHKLSDSES 160
Db 989 EPOEMD-----IDEILRLAETRENEVSTATDELL--SQFKVANFATMEDEEL-----EER 1038
QY 161 PNIFWSKLGGKPMWKPSPDTPQRNRKRVQVPE-----GSEASPTG 203
Db 1039 PHKDWDEII-----PEQRKKEVEERQKELEEIYMLPRISSTKKAQNTDS 1085
QY 204 DGNNAKKRKASDDVTPDPRVTPDDDERKASGKHMGALSPKVTITLOSSCKSSGTDG 263
Db 1086 DSDTESKROAQRSSASETED---SDDDKK-----PKRRGRPRSVRKDLVEG 1130
QY 264 TLIDGN-----DAGLYSMGSHSIGIPEDML--ASQDWGKIPDESQRRLHTVLKPKWAKLC 316
Db 1131 FTDAEIRRFIKAYKFE-----GLPLERLECLARD--AELVDKSVADL-----KRLG 1174
QY 317 QVLHLSDACTSMVGNFLEYVNIENRIYEPATTFQAFQIALSWIAALLVKQIILSHKESLV 376
Db 1175 ELIH--NSCVSAMQVEEQLKENASEGKGPGR--RGPTIKISGV-QVNVKSIHQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEDYIYVLSCKMSKSLFLEHTQGLQDFCGFNNSK 422
Db 1231 MLHKSIPVDPPEKKKYCLTCRVKAAHFDVGEWVEDDRLLLGLIYEHGYG--NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVDEKINTKSMNSSE-----DEECWTEKR 464
Db 1290 -----LKLTKILFVET--DKXPQQLQQLQTRADYLLKRLKGLKGGKAVGTGGEAKLKR 1342
QY 465 -----CSHYSTATRDIKTIISGIIKKYKQVQKLVQEHHEEK 501
Db 1343 KPRVKKENKVPRLKEEHGIELSSPRHSDNPSEGEVKDDGLEKSPMKKKQK--KKENK 1401
QY 502 MELLNNYAK-----KQKLETSKVEA-----AVIRITCSRTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKGKSKDKKEPKSGDAKSSSKRSQSPVHITAGSEPVPIGEDED 1461
QY 540 -----LKLIDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DLDQETFTSICKERMPVKALKQDLKDPKGLNVQSELEHTRNCLLKIGRIAECUKAYS 1521

```

QY 565 -----LEQMDVAKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DOEHLKWRNLWIFVSKTFEPDARKLHKYMAHKRSQEEB-----EQKKDGV 1572
QY 599 PTQSGNKHFF-----SGSSN-----ISQN 617
Db 1573 ----TGKKPFRPEASGSSRDSLISQS 1595

RESULT 8
QYRG05
ID Q7RG05 PRELIMINARY; PRT; 1268 AA.
AC Q7RG05;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 10.
GN Name=PY04546;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteza M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J.K., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RA "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.",
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01001392; EAA16423.1; -.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR InterPro; IPR000717; PCI.
KW Initiation factor.
SQ SEQUENCE 1268 AA; 153041 MW; DE3D62D55C47B91C CRC64;

Query Match 4.9%; Score 158.5; DB 2; Length 1268;
Best Local Similarity 20.2%; Pred. No. 2.3;
Matches 151; Conservative 110; Mismatches 253; Indels 233; Gaps 36;

QY 5 FTLFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCT-----VVEKALILARQNRON 59
Db 58 FCLYLEKLSLVKDLGLGHQYRI-----ICOHGNIASLGVITDFRDKAEK-VRLAKENMTN 112
QY 60 KAVENLNRLSLTHALLMWGASYLFDKLDHFHSSPTDPSGVSF-EQSIMDGVIFEFSSILSS 118
Db 113 K-----DKI-----EQESNVDFTEKILMNSLDWEITGKYER 144
QY 119 KGEENEVKLC-----LLLEAKHA-----QGTY-----SSDSTLFGEDHKLSD-----EESPN 162
Db 145 K--LQNAVYKICMETYKMWILEILRATPKLEKAVHDTAKALLFCENKRLTEPKLSDLRL 202
QY 163 IFWKLKLGKNP-----MWKYPSTPQ--RNRKEVOYF 193
Db 203 NHYNLILRKGKHPYOSLKLIEYHLETKIMQLETACELGHWKEASIASDIYNLNMHEYP 262
QY 194 -----EGSASPKTGGNAKKRKKASDVTDPVDDDERKASGDHMGALSPK 248
Db 263 YKTLKSGAMESDNNWADQNDKSKSSD-----KOEKDTNEATTSGGVASTS 310
QY 249 VITLQSSCK-----SSGTDGTLGDNDAFGLYSMGSHIS----- 281

```

```

Db 311 ATASQNKQENLKKWIATFYEKLADILFYESLPHGLAWLKYCYPHILNLYGIEISEERKK 370
QY 282 -----GIPE-DMLASODWGIKIPDESORRLHTVLKPKWAKLCQVLHLSDACTSMV 329
Db 371 FICTKAVLAVLSIPRIGNKKNEDIYAKI-FAEHKGSQQLGHTSVPKESLKGK-----LVKQILSHKESL 375
QY 330 GNFLEYVIEH-HRIYEEPPATTFQAFQIALSWIAAL-----LVKQILSHKESL 375
Db 426 RNILYADENAKQLYSLVENKFTPLSLCLECEVLLKELEKTEHNTYINKIVEIPHKLIL 485
QY 376 VRANSELAFKCSRVEDY-IYSILS-----CMKSLFLEHTQGLQFDC-----F 417
Db 486 -----QLSKVYNYISIDYFIENICSEHFISWMDAETMLVDLYVQKELEMRIDLTKRAIYF 540
QY 418 GTNSKQSVVSTKLVNESL-----SGATVRDEKINTKSMRNSSEDEECMTE 462
Db 541 G--DRENVRSKLIKDLSLINMYNLYGLRMINESISENEVENTNTKLLNYHEHFNLLD 598
QY 463 KRCSHYST-ATRDIEKTISG-----IKKKYKQVQKLVQEHHEKQKMLNMYADKK 512
Db 599 KGNKHISNATADIEDIVENEIIDDKPIDDEKILKIVDKIDEEHH--KLQLLSEHNKK 656
QY 513 QK--LETSSVEAAVIRITCSRTSTQVGLKLLDHNRYERKPFDEIKSEKNECLKSLEQMDH 570
Db 657 RKEMLRKQEQQAOLKIRMEK-----RLLEKLEKEKEKEELAKGKELRIKEE--K 706
QY 571 VAKKKLAEDACWINRIKSWAAKLKVC 597
Db 707 IKKSEAAEQ--MLKEIK-----KLC 725

RESULT 9
QYRG05
ID Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein PF08_0127.
GN Name=PF08_0127;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Bartell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51332.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Query Match 4.9%; Score 158; DB 2; Length 1245;
Best Local Similarity 18.1%; Pred. No. 2.4;
Matches 125; Conservative 112; Mismatches 258; Indels 196; Gaps 27;

QY 12 LNPSHDVHKVEKIKIESCSERTKIFRLY-----SVCTVEEKALILARQNRQNKAVENLN 66
Db 1 MWYDEDIKRRKKTKTEKLNKTESLSLYESNNKILSIKKNKLLKDKKKNKIKQMKN-- 58
QY 67 RSLTHALLMWGASYLFDKLD-----HPSSETPDSPGVSPQSIMD----- 106
Db 59 -----KYDKDGTIKHKGNKNISSDDKDFDEDITNLYMKSKSKCKMDTMYK 104
QY 107 -----QVIHEFSSILSSKGEENEVKLCILLEAKHAQGYSSDSTLF 148
Db 105 KKKKKKKKNQYKKNYIGVMDDEIIDD--NEERIIENNVNNDVHPYINDSK---- 159
QY 149 GEDHIKLSDEESPNFWSKLLGGKQPMWKYPSDTFQNRKRVQYFEGSBASPKTGGGNA 208
Db 160 -----SDEEYSSVNFMKFLKCN-----DNREVNKKKKKKKSTDGNHNDHND 206

```

```

QY 209 KKKKASDDVTPRVTPPPVDDDERKASGKHMHGALESKPVIITLQSSCKSSGTDGTLGN 268
DB 207 NDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDN 247
QY 269 DAFGLYSMGSHISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLC-----QVLHLSDA 324
DB 248 DDNIIKKLVH-SGDQDDHVDSSSVQDSYQ---YYALKDDMYKLSNENKIEIPADL 303
QY 325 CTSVMGNF-----LEYVIENHR-----IYEEPAITFQAFQIALSWIAALLVK 366
DB 304 MYGN-NNFFQGLQKLDLF-LENEQNKKKGKRLSPTEB-----QKFNTTLGVVQNVKLL 356
QY 367 QILSHKESLVRANSELAF-----KCSRVEVDYIYSILSCMKSLFLEHTQGLQDFDCGTNS 421
DB 357 NQNKAYVNSKVFGRQKRGNGQSGDFVYKRL---VGRFLRKRKEKIQPEMYDGH 413
QY 422 KQSW---STKLVSLSGATVDEKINTKSMNSSEDEECMTKCRSHYSTATRIEKT 478
DB 414 ESGYIDNDDNIIINNI---LKRHKVDQTHNGYGEQNEVYKK-----EKK 457
QY 479 ISGIIKKYKQVQKLVQEHHEEK-KMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQV 537
DB 458 I--IKREHKKEEDFTIYNPEEIKMNL-----KSKMLHISED-----493
QY 538 GDLKLLDHNVERFDEIKS-EKNECLKSLEQMDHVAKKLAEDACWINRIKSWAAKLVK 596
DB 494 ----LKNKEKRDLEKTAQKLLLSIENRKNYKKIKSKSYRKYLRIRKEKBEEKI 549
QY 597 CVPIQSGN-----NKHFGSSGNISON 617
DB 550 MKKLYSEHPDLVKDLNMYEYAEKRNLLNN 580

RESULT 10
Q9JK25 PRELIMINARY; PRT; 1320 AA.
AC Q9JK25;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CLIP-170.
GN Name=cynl;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RX MEDLINE=21185938; PubMed=11290329; DOI=10.1016/S0092-8674(01)00288-4;
RA Akhmanova A., Hoogenraad C.C., Drabek K., Stepanova T., Dorthland B.,
RA Verkerk T., Vermeulen W., Burgering B.M., de Zeeuw C.I., Grosveld F.,
RA Galjar N.;
RT "Claps are CLIP-115 and -170 associating proteins involved in the
RT regional regulation of microtubule dynamics in motile fibroblasts.";
RL Cell 104:923-935(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Hippocampus;
RA Galjar N.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237670; CAB92974.1; -
DR HSSP; Q20728; 1LPL.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000938; CAP-gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_GLY; 2.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS0245; CAP_GLY_2; 2.
SQ SEQUENCE 1320 AA; 148563 MW; 87BE0F0463D20E9D CRC64;

```

Query Match

4.9%; Score 157; DB 2; Length 1320;

```

Best Local Similarity 20.6%; Pred. No. 3;
Matches 124; Conservative 98; Mismatches 225; Indels 156; Gaps 25;

QY 19 KHVEKIKIESCSRTKIFRLYSVCTVEEKALILARQNMQRONKAVENLNRLTHALLMWA 78
DB 484 RELEDFRVATVSEKSRIMELEKDLARVQEAELRRLESSKPPGVDNLSL---LLEQEI 540
QY 79 SYLFDKLDHMFHSETPDGVSFEQSGMDGVHIFHSSILSKSGEE-----NEVKLCLLLEA 134
DB 541 SALQBLEVTHTDHQ-----NEVTSLKOHFGTREENFQKEIKHAHAATE 584
QY 135 KHAQGYSSDSTLFGEDHIKLSDESPNIFWMLGKGNPMWKPVPDTPQNRKRQVYFE 194
DB 585 KLSKENESLRSLU---DH---ANKNSDVI-----ALWKSULETATASHQ--QAME 627
QY 195 GSASPKTDGGNAKRRKKASDDVTPRV-----TDPVDDDERKASGKHMHGALESP 247
DB 628 ELKVSFSKGIQDTSASFABELKQIERLDYQHEIESLQSKDSERSAHAK-----MESM 683
QY 248 KV-----ITLQSSCKSSGTDGTLGDNDAFGLYSMGSHISGIPEDMLASQDWGKIPDESQR 303
DB 684 KAKLMKIIKEKEDSLEAVKARLDTABDQHLVEMEMLSKLQEAIEIKKEKFASEEA--- 740
QY 304 LHTVLPKMAKLCQVHLSDACTSMVGNFLEYVIENHRIYEEPAITFQAFQIALSWIAAL 363
DB 741 --VSTQTSQDVTNKLHQKEEQNMLUSLEKLEN-----LTDMEA-----780
QY 364 LVQKILSHKESLVRANSELAFKCSRVEVDYIYSILSCMKSLFLEHTQGLQDFGNTSKQ 423
DB 781 KFKEDEREDQLVKAKEKL-----END-----IAEIMK-----MSGDNSSQ 816
QY 424 SVVSTKLVNE-SLSGATVDEKIN-TKSMRNSSEDEECMTKCRSHYSTATRIEKTISG 481
DB 817 ---LTKNDLRLKRSVEELQKLTKANENASL-----LOKSIGE 854
QY 482 IKKYYKQVQKLVQEHHEEKMELLNMYADKKLETSKSVAAVIRITCSRTSTQVGDLC 541
DB 855 VTLKAEQSQAQAKGHEEKLEKLEKMETSK-----HYQCQDLK 899
QY 542 LLDHNYERKFDTEIKSEKNECLSEQM-HDVAKKLAEDACWINR-----IKSWAA 592
DB 900 A---KYEKASSETKIKHEEILQNFQKMLVDTEDKLKAAQEA---NRDLQMDMELKSLQAD 953
QY 593 KKK 595
DB 954 KAK 956

RESULT 11
Q7SHZ4 PRELIMINARY; PRT; 4007 AA.
AC Q7SHZ4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU00658.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Raemussen C., Metzberg R.L., Perkins D.D., Kroken S.,

```



```

Db 710 SDHTVQEDAVKLSKQNTIVANWTPRNEKLFYLLHQTPARDFEIIHHDGDKLNRVKEL 769
Qy 333 -----LEYVIENHRIYEEBPAITFOAFQIALSALIAALLVQILSHKESLVRANSELAFKC 386
Db 770 GESVESLELEKNKSTP-----SNLEAKLAEMTWSHKKEIENTQAB-----811
Qy 387 SRVEVDYIYILSKMSKLFLEHTQGLQDFGNGNSKQSVVSTKLVRNLSGATVRDEKIN 446
Db 812 -----CIKMSVEFE--LMTDSMSRQSKQIESKDRREIEELKAKLEKQNIH 856
Qy 447 TKSMRNSSEDEECWTEKRCCHSYSTATDIEKTSIGIKKKYKQVQKLQV-----495
Db 857 EKALRNDPYSEE-----YKTLTAETR--AELEKPKQIEVITTKAVECKKDEAPA 905
Qy 496 -----EHEKEMELINMYADKKQKLET--SKSVEAAVI-----RITCS 531
Db 906 ROEKTLEIENRVLSSESKKLEAWNR--EKEQLESILRQMPGEVILDSFNALKES 962
Qy 532 RTSTQVGD-----KLJDNHYVERKPEI-KSEK-NECLSKSLQEMHVDVAKKK 575
Db 963 QPSAQIKDPAAIRSRMERSTGKCLDTNKVLLIDNILKVHKLQQLVNEREQYRDVKSQ 1022
Qy 576 LADEACWINRIKSWAAKLVKCVPIQSGNNKHFGSGS 611
Db 1023 PGDQV--LQFLENSTPTVPDMEAYFNELTNFGSGS 1056

RESULT 13
Q6BHF8
ID Q6BHF8 PRELIMINARY; PRT; 905 AA.
AC Q6BHF8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA2768|PF17011 Candida albicans |PF17011.
GN ORFNames=DEHA020119;
OS Debaryomyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=284592;
RN [1]_TaxID=284592;
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marcq C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistrame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolaki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Peillen S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swemene D., Tokia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382139; CAG90870.1; -.
DR InterPro; IPR002017; Spectrin.
SQ SEQUENCE 905 AA; 103103 MW; F7B4B8E0D525F0C4 CRC64;

Query Match 4.8%; Score 155; DB 2; Length 905;
Best Local Similarity 19.0%; Pred. No. 2.4;
Matches 131; Conservative 107; Mismatches 239; Indels 214; Gaps 29;

```

```

Qy 52 ARQNMQNKAVENLNRLSLTHALLMWGASYLFD-----KLDHFHSSTPD 95
Db 48 AKQDASRNTVHNETPGLTATPLTKKAEPYVKNRIPLTSPRATNPDLNLSLKSKTK- 106
Qy 96 SGVSPQSIQMDGVTHFESSILSSKGGEEVLCLEAKHAQGYSSDSTLFGEDHKL 155
Db 107 -----LDARLYELKS--SSRHGSPN-----KENHNSGKSTQASQLSDVNSSI 149
Qy 156 SDESPNIFWSKLLGKGNPMWKYPS--DTPQRNRKRVQVPEGEASPKTGDGNAKRRK 213
Db 150 SH-----KSPFESSPKIDTIK-QINSSFAKNVKSPPKSPSKADKRS- 193
Qy 214 ASDVDTPRVTPVDDDERKAGKDHMGALSPKVTITQSSCKSS-----GTGD 263
Db 194 -----IEDDSVIENTSRKIA-----KVTIANESEVSHKSDDEHLIQIDQ 233
Qy 264 TLDGNDAPGLYSGSHISGIPEDMLASQDW-GKIPDESQRRLHTVLKP-----KWAKLCQV 318
Db 234 VFNKEDLVGRFSGGAPAKLHEPTVIDEDSDGDIYNNVNTSHTMTSKMKGNESTQLSRI 293
Qy 319 LHLSDAC-----TSMVGNFLEYVIENHRIYEEBPAITFOAFQIALSWIAALLVKQILS 370
Db 294 IKITDEVYVTPQRYTEPVAKSP--INQHRIKENVINNTESK-----TISPLKN 343
Qy 371 HKESLVRANSELAFKCSRYE-----VDYIYISILSCMKSLF-LEHTQGLQF 414
Db 344 HMLSVDNOKENHVSNNRMEDENSDGFNELEDEPTINFLLSPNS--KPVFLSHDKKVKQD 401
Qy 415 DCFG-----TNSKQSVSTKLVRNLSGAT-----VRDEKIN-----446
Db 402 DHFKEVANLEEVINNKNQELK--FSELSATNNKFLIYDQIKELKQDKKLIANENL 458
Qy 447 -----TKSMR-----NSSEDEECWTEKRCCHSYSTATDIEKTI-----479
Db 459 LLIQLKHNERELASMTKALRIKENTVTOLESRLSKSKSYESTANELESVIEGNTLREQ 518
Qy 480 -----SGIKKKYKQVQKLQVQHEEKKME--LLNMYADKKQKLETSKSVEAAV 525
Db 519 IKRLEKSVEDKSIQKESQIKILNDEIKEDTETSSLTANDIDYNIKNVENLLSEKEEL 578
Qy 526 IRTCSRTSTQVGDLLKLDHNYVERKPEIKSEKNECLSKLEQMHVDVAKKLADEACWIN 585
Db 579 L-----TETGRGRENNGLE-----EINSKQOELLELDKLENLAKOKIMTLE---N 622
Qy 586 RIKSWAAKLVKCVPIQSGNNKHFGSGSSNISQ 616
Db 623 TLDSKTQBIK--QVWTEKNELLTKIDNLTE 650

RESULT 14
Q676A5
ID Q676A5 PRELIMINARY; PRT; 1033 AA.
AC Q676A5;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Serine/threonine protein kinase (Fragment).
GN ORFNames=002-44;
OS Oikopleura dioica.
OC Eukaryota; Metazoa; Chordata; Urochordata; Appendicularia;
OC Oikopleuridae; Oikopleura.
OX NCBI_TaxID=34765;
RN [1]_
RP SEQUENCE FROM N.A.
RX PubMed=15343333; DOI=10.1038/nature02709;
RA Hansen A., Flaatt M., Weissenbach J., Lehrach H., Wincker P.,
RA Reinhardt R., Chourrout D.;
RT "Hox cluster disintegration with persistent anteroposterior order of
expression in Oikopleura dioica.";
RL Nature 431:67-71 (2004).
DR EMBL; AY49458; AAS21362.1; -.

```

```
DR GO: GO:0016301; F:kinase activity; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR InterPro: IPR000194; Atpase a/bcentre.
DR InterPro: IPR011009; Kinase-like.
DR InterPro: IPR000719; Prot.kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Prot.kinase; 1.
DR SMART: SM00220; S_TK; 1.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00152; ATPASE ALPHA BETA; UNKNOWN_1.
DR PROSITE: PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00108; PROTEIN KINASE-ST; 1.
KW Kinase; Serine/threonine-protein_kinase.
FT NON_TER 1
SQ SEQUENCE 1033 AA; 118986 MW; 9C1D5559307395F7 CRC64;

Query Match          4.8%; Score 154.5; DB 2; Length 1033;
Best Local Similarity 18.9%; Pred. No. 3;
Matches 134; Conservative 108; Mismatches 241; Indels 225; Gaps 27;

QY 26 IESCERTKIFRLYSVCTVEKALILARQNMQRNKAVENTLRSLSLTHALLMWSGLYFDKL 85
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 VSKSTRKNSVGSMSLSSEENSEV---QPSKYERKAIHQKFNVSSEA----- 278
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 DHFSESETP-----DSGVFEQSDMGVTHE---FSSILSKSGEENEVKLCILLEA 134
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 ---NSSESPPTVPFTSSSGVS---EIGENGHSNGVFSNSDTSQTSKSENPIL----- 327
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 KHAQGYSSDSTLFGEDHKLSDSESPNIFMSKLLQ---GKNPMWYKPDTPQNRKRQV 191
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 -----TESLDST---EEIQVDKNTVRVKGGLFGKNASKLAKQLSQSDTIVHNSKSV- 377
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 YFEGSEASPKTGGNAKRRKAS--DDVTDP-----RVTDPVDDDERKASGKHMGAL 244
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 378 -----LPKANSVDVEVLGKMSGPDVIAPHSPVKQNTNPLNSEKSKSPNTHTKSM 430
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 ESPKVTITLSSCKSGTDLQGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDESQRL 304
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 NSKMQDHLPAS---SSDSKGTWFG-----TSKSKMAAPPQSAEGI 467
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 HTVLKPKMAKLQVLHLSDACTSWGNFLEYVIENHRI-----Y 343
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 468 GEMFKEEIKK-----EDEALEMLDNLVLLSEKSHSVVNYVSAGKTPVQVGAARESY 520
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 EEPATTFOA-----FQIALSWIAALLVKQILSHKESLVANS 380
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 KNATTETEAKTNKPVILTSTENSVEDETEILYEKKVPTKSEMKKKTTLTHEEKLKRQK 580
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 ELAPKCSRVEVD-----YIYSLSCMKSLF-----LEH 408
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 581 TLK-KTRKFIVDGEEVITTSINIPVRRFRFKYKMYVTSDHDVDEQSKRIFRQELQD 639
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 409 TQGLQFDCFGTNSKQSVSTKLWNESLGSATVRDEKINTKSMRNSSEDEECETKRCSHY 468
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 640 LRKLQIE-----EQKMANLNWRLA-AQTEQLKTHDRETELAKIYQTKIDSLE 688
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 469 SPATRDIEKTIIGIKKKYKQVQKLVQE-----HEKQWELLNMYADKKQ 513
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 689 KLQKHKIEKWSQAQEKQEMISTLKREQNKELIAFQQQSHASHSELKRVTAQFSDRKA 748
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 514 KLETSKSVAAVIRITCSRTSTQVGLKLLDHYEKFDS--IKSEKNECLK-SLEQMDH 570
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 749 RTETURA-----RKQELNSHOKYKEGAFKAFQEDKLVESVSKFSD 788
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 571 VAKKKLAEDACWIN-RIKSWAAKLKVCVPISQGNKNKHFSGSNISQN 617
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 789 NQKTVAEETERRFLDQKQESLLAREKAIWDLK---EKHFQEKHQVAKS 833
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```


DR ENBL; X82834; CAAS8041.1; --
DR ENBL; U31906; AAC51791.1; --
DR ENBL; X76942; CAAS4261.1; --
DR PDB; 1RA4; X-ray; E/F/G/H=2172-2222.
DR PDB; 1UPT; X-ray; B/D/F/H=2169-2228.
DR Genew; HGNC:4427; GOLGA4.
DR MIM; 603509; --
DR MIM; 270150; --
DR GO; GO:0005802; C:Golgi trans face; TAS.
DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR InterPro; IPR000237; GRIP.
DR Pfam; PF01465; GRIP; 1.
DR PROSITE; PS50913; GRIP; 1.
KW 3D-structure; Alternative splicing; Antigen; Coiled coil; Golgi stack.
FT DOMAIN 133 2185 Coiled coil (potential).
FT DOMAIN 252 2096 Glu-rich.
FT DOMAIN 2168 2215 GRIP.
FT VARSPIC 2154 2185 TPYKGNLVHTDVSLEPTEPEVLRKVLPEY -> HLTKV
FT AICTIRMSHLENLPNLSICEKFLSI (in isoform
2).
FT /FTid=VSP_004272.
FT Missing (in isoform 2).
FT /FTid=VSP_004273.
FT Missing (in isoform 3).
FT /FTid=VSP_004274.
FT FTSPRGIF -> SWLRSS (in isoform 4).
FT /FTid=VSP_004275.
FT R -> K (in Ref. 3).
FT CONFLICT 188 188 Y -> H (in Ref. 3).
FT CONFLICT 220 220 T -> A (in Ref. 3).
FT CONFLICT 276 276 K -> E (in Ref. 3).
FT CONFLICT 584 584 T -> A (in Ref. 3).
FT CONFLICT 628 628 K -> E (in Ref. 3).
FT CONFLICT 630 630 K -> N (in Ref. 3).
FT CONFLICT 682 682
FT SSEQUENCE 2230 AA; 261139 MW; 3BE733DB1EA86134 CRC64;

Query Match 4.8%; Score 153.5; DB 1; Length 2230;
Best Local Similarity 19.7%; Pred. No. 9.2;
Matches 135; Conservative 120; Mismatches 268; Indels 161; Gaps 31;

QY 13 NPSHDVKHVEKI-KIESCS-----RTKIFRLYSVCTVEEKALILARONMQNKAIVEN 64
DB 1552 NQKDIHEKELVQLQHFQELGEEKNRVK-----EAEKILITLENQVSMKALET 1603

QY 65 LNRSLTHALMWGA-----SYLFDKLDHFSSETPDGSGVFEQSIMDGVHIEFSSILSSK 119
DB 1604 KKELEHVNLSVSKSEELKALEDRLSESAKLAELKRAEQKIA-AIKKQLLSQMEEK 1662

QY 120 GGEENEVKLCILLEAKHAQTYSDSTLFGEDHIKLS-D-BESNIFWSKLLGGKNPMWKY 178
DB 1663 -----EEQYKGTESHLSEL-----NTKLQEREREVHILEEKLKSVESSQ--- 1702

QY 179 PSDT---PQRNRKVQYFEGSEASPTGDCGNKAKRKKASDDVTDPRVTDPPV-----D 229
DB 1703 -SETLIVPSAKNVAAITEQEADSQ---GCQKTYEEKISVLQRLNTEKEKLLQORVGQ 1757

QY 230 DDERKASGDHMGALSPKIVITLQSSCKSGCTDGLD-----NDAFLGYSMGSHI 280
DB 1758 EKETVSSHEFEMRCQYQERLIKLEHAEAKQHEHQDSMIGHLEELKEKNKYSLI-VAQHV 1816

QY 281 --SGIPEDMLASQDWGKIPDESQRRLHTVLKPKWAKLCQVLHLS-----DACTSMVGNFLE 334
DB 1817 EKEGGKNIQAKQNLNENFDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864

QY 335 YVIEHRI-YEEPATTPQAFQIALSWI-----AALLVKQILSHKES 374
DB 1865 RQKEVHRVMEETSKYEKLQ-ALQQMDGRNKPTELEENTEKSKSHLVQPKLLSMEA 1923

QY 375 LVANSLELAFKCSREVD-----YIYSILSCMKSLFLEHTQGLQFDCFGTNSKQSVST 428
DB 1924 ---QHNDLEFLAGAREKQKLGKEIVRLQKDLRLMLRKEHQOELE-----ILKK 1969

QY 429 KLVNESLSGATVRDEKINTK-----SMRNSSEDEECMTEKRCSHYSTATRDIEKTIIGIKK 484

Db 1970 EYDQE-----REEKIKOEQEDLEKHNSTLQLMREFN-TOLAQKEQELENWIKETIN 2021
QY 485 KYKKQVQKLVQEHHEEKMELLNMYADKKQKLETSKSVEAAVIRITCSRTSTQVGDCLK-- 542
Db 2022 KAQEVEAEELLESQEEETNQLLKIAEKDDDLKRTAKRYEIEILDAREEEMTAKVRDLQTL 2081
QY 543 --LDHNYERKFDIEKSEKNECLKSLEQMDHVAKK-----KLAEDAEACHINRIKSWAAK 593
Db 2082 EELQKKYQOKLEQENPGNDNVTIMBLQTLAQKTTL1SDSKLKEQE--FREQIHNLDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
Db 2140 LKKYKKNVATTVGTPTYKGNLYH 2163

Search completed: July 12, 2005, 01:01:44
Job time : 65.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 13:13:50 ; Search time 63 Seconds
(without alignments)
3812.352 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270

Perfect score: 3216
Sequence: 1 RADAFILFSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseq19808:*

2: Geneseq19908:*

3: Geneseq20008:*

4: Geneseq20018:*

5: Geneseq20028:*

6: Geneseq20038:*

7: Geneseq20038b:*

8: Geneseq20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3213	99.9	2001	4	AAB20063 Arabidops
2	3210	99.8	2001	4	AAB20062 Arabidops
3	163.5	5.1	1465	8	ADN22447 Bacterial
4	162.5	5.1	842	8	ADQ66710 Novel hum
5	162.5	5.1	1739	6	AAB36105 Human chr
6	162.5	5.1	1739	6	AAB36106 Human chr
7	162.5	5.1	1739	8	ADK60208 Angiogene
8	162.5	5.1	1739	8	ADL12347 Human ste
9	162.5	5.1	1739	8	ADK60509 Angiogene
10	162.5	5.1	1739	8	ADP73132 Angiogene
11	161	5.0	1015	8	ADN20557 Bacterial
12	156	4.9	954	4	AAB14615 Novel bon
13	154.5	4.8	2228	7	ABR61599 Human gol
14	154.5	4.8	2230	6	ABU07445 Protein d
15	154.5	4.8	2230	7	ABR61600 Human gol
16	154.5	4.8	2250	7	ABR61601 Human gol
17	154.5	4.8	2252	7	ABR61602 Human gol
18	151.5	4.7	1931	4	ABR61012 Drosophll
19	150	4.7	1948	4	ABG21233 Novel hum
20	148.5	4.6	823	4	ABU53074 Intracell
21	148.5	4.6	875	4	ABU53078 Intracell
22	148	4.6	1455	7	ADK80783 Microsate
23	148	4.6	1881	5	ABP73809 Candida a
24	148	4.6	3187	7	ADK56031 Rat Prote
25	148	4.6	3187	7	ADK56035 Rat Prote

26	147	4.6	962	3	AAB18211 Plasmodiu
27	147	4.6	3225	7	ADJ68448 Human hea
28	146.5	4.6	1427	2	AAR10534 Human 160
29	145.5	4.5	1647	5	ABP73735 Candida a
30	145.5	4.5	1959	5	ABJ10604 Human nov
31	145	4.5	1130	8	ADN24098 Bacterial
32	145	4.5	3032	8	ADT71250 Human Dro
33	144.5	4.5	1392	2	AAY06999 Restin pr
34	143	4.4	1379	7	ADB79765 Rat Rho k
35	143	4.4	1960	4	AAM78854 Human pro
36	143	4.4	1960	7	ADB70370 PAC 6802
37	143	4.4	1960	7	ADDA7946 Human pro
38	143	4.4	1960	8	ADJ37149 Human mal
39	143	4.4	1960	8	ABM82345 Tumour-as
40	143	4.4	1960	8	ADN60283 Human nos
41	143	4.4	2143	4	ABG01716 Novel hum
42	142.5	4.4	1427	8	ADP56353 Human PRO
43	142	4.4	1738	8	ABM83885 Human dia
44	140.5	4.4	1788	4	AAM40467 Human pol
45	140.5	4.4	1963	4	AAM79838 Human pro

ALIGNMENTS

RESULT 1

AAB20063

ID AAB20063 standard; protein; 2001 AA.

XX AAB20063;

XX 23-APR-2001 (first entry)

XX Arabidopsis thaliana silencing gene-encoded protein.

XX Gene silencing; silencing gene; MOM.

XX Arabidopsis thaliana.

XX Location/Qualifiers

FT Region 177..350

FT Peptide /note= "internal repeat region"

FT Binding-site 362..367

FT Region /note= "nuclear localization motif"

FT Region /note= "ATP/GTP-binding motif"

FT Region /note= "region of homology to ATPase/helicase family"

FT Misc-difference 705

FT Peptide /note= "Met in ecotype Zurich"

FT Peptide 832..838

FT Peptide /note= "nuclear localization motif"

FT Domain 858..862

FT Domain /note= "nuclear localization motif"

FT Misc-difference 1219

FT Region /note= "predicted membrane-spanning domain"

FT Region 1462..1672

FT Region /note= "internal repeat region"

FT Region 1848..1894

FT Domain /note= "internal repeat region"

FT Domain 1899..1941

XX /note= "actin-binding domain"

XX WO200100801-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.

XX 23-JUN-1999; 99GB-00014623.

```
XX (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
DR WPI; 2001-137952/14.
DR N-PSDB; AAA89385.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
XX Claim 8; Page; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana protein MOM, which
CC controls gene silencing, particularly gene silencing. It is encoded by a
CC novel silencer gene (see AAA89385) of A. thaliana ecotype Columbia. The
CC protein exhibits partial similarity with ATPase/helicase proteins of the
CC SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
CC compared with the protein (see AAB20062) encoded by the silencer gene of
CC Arabidopsis thaliana ecotype Zurich. Gene silencing is useful as a
CC molecular tool for regulating gene expression. Note: the present sequence
CC is not shown in the specification but is derived from the A. thaliana
CC mutant silencer gene protein sequence given in AAB20062
XX
SQ Sequence 2001 AA;
Query Match 99.9%; Score 3213; DB 4; Length 2001;
Best Local Similarity 99.8%; Pred. No. 7.5e-266;
Matches 620; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 RADAFILFGSSLPNSHDVXVVEKIKIESCERTKIFRLYSVCTVEBKALILARQNKQNK 60
Db 650 RADAFILFGSSLPNSHDVXVVEKIKIESCERTKIFRLYSVCTVEBKALILARQNKQNK 709
QY 61 AVENLNRSLTHALLMGWASYLFDKLDHFHSSSTPDSCVSGVFEQIMDGVHFEFSSILSSKG 120
Db 710 AVENLNRSLTHALLMGWASYLFDKLDHFHSSSTPDSCVSGVFEQIMDGVHFEFSSILSSKG 769
QY 121 GEENEVKLCLEAKHAQTYSSDSTLFGEDHIKLSDESPNIFWSKLGKGNPMWKYPS 180
Db 770 GEENEVKLCLEAKHAQTYSSDSTLFGEDHIKLSDESPNIFWSKLGKGNPMWKYPS 829
QY 181 DTPQRNRKVQVFESEASPKTGDGNAKRRKASDDVTDPRVTPDPVDDDERKASGKD 240
Db 830 DTPQRNRKVQVFESEASPKTGDGNAKRRKASDDVTDPRVTPDPVDDDERKASGKD 889
QY 241 MGALESPPKVIITLQSCCKSGTDTLGDNDAPGLYSMGSHISGIPEDMLASQDWGKIPDES 300
Db 890 MGALESPPKVIITLQSCCKSGTDTLGDNDAPGLYSMGSHISGIPEDMLASQDWGKIPDES 949
QY 301 QRLRLTVLKPMAKLCQVLHLSDACTSMVGNFLEYVYENHRIYEEPPATTFQAFQIALSWI 360
Db 950 QRLRLTVLKPMAKLCQVLHLSDACTSMVGNFLEYVYENHRIYEEPPATTFQAFQIALSWI 1009
QY 361 AALLVKQILSHKESLVANSELAFCRSRVEVDYIYSILSCMSKSLFLEHTQGLQFCFGTN 420
Db 1010 AALLVKQILSHKESLVANSELAFCRSRVEVDYIYSILSCMSKSLFLEHTQGLQFCFGTN 1069
QY 421 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 480
Db 1070 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTIIS 1129
QY 481 GIKKKYKQVQKLVQHEEKKNMELLNMYADKKQLETSKVAARITCSRTSTQVGD 540
Db 1130 GIKKKYKQVQKLVQHEEKKNMELLNMYADKKQLETSKVAARITCSRTSTQVGD 1189
QY 541 KLLDHNRYERKFEIKSEKNECLKSLQEMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 600
Db 1190 KLLDHNRYERKFEIKSEKNECLKSLQEMHEVAKKLAEDACWINRIKSWAAKLKVCVPI 1249
QY 601 QSGNNKHFGSSNISQNAADV 621
|||||
```

```
Db 1250 QSGNNKHFGSSNISQNAADV 1270
RESULT 2
AAB20062
ID AAB20062 standard; protein; 2001 AA.
XX
AC AAB20062;
XX
DT 04-NOV-2004 (revised)
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene-encoded protein.
XX
KW Gene silencing; silencing gene; MOM.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Region 177..350
FT /note= "internal repeat region"
FT Peptide 362..367
FT /note= "nuclear localization motif"
FT Binding-site 460..467
FT /note= "ATP/GTP-binding motif"
FT Region 479..719
FT /note= "region of homology to ATPase/helicase family
FT SWI2/SNF proteins"
FT Misc-difference 705
FT /note= "Lys in ecotype Columbia"
FT Peptide 832..838
FT /note= "nuclear localization motif"
FT Peptide 858..862
FT /note= "nuclear localization motif"
FT Domain 995..1015
FT /note= "predicted membrane-spanning domain"
FT Misc-difference 1219
FT /note= "Asp in ecotype Columbia"
FT Region 1462..1672
FT /note= "internal repeat region"
FT Region 1848..1894
FT /note= "internal repeat region"
FT Domain 1899..1941
FT /note= "actin-binding domain"
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-EP005761.
XX
XX 23-JUN-1999; 99GB-00014623.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
WPI; 2001-137952/14.
XX N-PSDB; AAA89353, AAA89354.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
XX silencing of plant genes.
XX
XX Claim 1; Page 32-39; 48pp; English.
XX
XX The present sequence is that of Arabidopsis thaliana protein MOM, which
XX controls gene silencing, particularly gene silencing. It is encoded by a
XX novel silencer gene (see AAA89353) of A. thaliana ecotype Zurich. The
XX protein exhibits partial similarity with ATPase/helicase proteins of the
XX SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
XX compared with the protein (see AAB20063) encoded by the silencer gene of
XX wild-type A. thaliana ecotype Columbia. Gene silencing is useful as a
```

CC molecular tool for regulating gene expression
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX Sequence 2001 AA;
QY Query Match 99.8%; Score 3210; DB 4; Length 2001;
Best Local Similarity 99.8%; Pred. No. 1.4e-265;
Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 RADAFILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKQK 60
650 RADAFILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKQK 709
61 AVENLNRLTHALLMWSGASYLFDKLDHFSHSETPDGSGVSPFQSDGIMDGVHEFSSILSSKG 120
710 AVENLNRLTHALLMWSGASYLFDKLDHFSHSETPDGSGVSPFQSDGIMDGVHEFSSILSSKG 769
121 GEENEVKLCILLAKHAQGYSSDSTLFGEDHIKLSDESPNIFWSKLLGGKPKMVKYPS 180
770 GEENEVKLCILLAKHAQGYSSDSTLFGEDHIKLSDESPNIFWSKLLGGKPKMVKYPS 829
181 DTPQNRKRQVYEGSEASPKTGGNAKRRKASDDVTPRVTDPVDDDERKASGKH 240
830 DTPQNRKRQVYEGSEASPKTGGNAKRRKASDDVTPRVTDPVDDDERKASGKH 889
241 MGALESPKVTITLOSSCKSSGTGTLGNDAGLYSMGSHISGIPEDMLASQDWGKIPDES 300
890 MGALESPKVTITLOSSCKSSGTGTLGNDAGLYSMGSHISGIPEDMLASQDWGKIPDES 949
301 ORLHTVLKPKMAKLCQVLHLSACTSMWGNFLEYVNIENRIYEEPAATTQAFQIALSWI 360
950 ORLHTVLKPKMAKLCQVLHLSACTSMWGNFLEYVNIENRIYEEPAATTQAFQIALSWI 1009
361 AALLVQQLSHKSLVRANSELAFKCSRVVDVYIYSLSCWKSLEHTQGLQDFCGTN 420
1010 AALLVQQLSHKSLVRANSELAFKCSRVVDVYIYSLSCWKSLEHTQGLQDFCGTN 1069
421 SKQSVSTKLAVNESLGSATVDEKINTKSMRNSSEDEECMTEKRCSHYSTATRDIEKTI 480
1070 SKQSVSTKLAVNESLGSATVDEKINTKSMRNSSEDEECMTEKRCSHYSTATRDIEKTI 1129
481 GIKKKYKQVQLVQHEEKKMELLNMYADKKOKLETSKSVAAVIRITCSRTSTQVGD 540
1130 GIKKKYKQVQLVQHEEKKMELLNMYADKKOKLETSKSVAAVIRITCSRTSTQVGD 1189
541 KLIDHNYERKFDIKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLVKCVPI 600
1190 KLIDHNYERKFDIKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLVKCVPI 1249
601 QSGNNKHFSGSSNISQNPV 621
1250 QSGNNKHFSGSSNISQNPV 1270

RESULT 3
ADN22447
ID ADN22447 standard; protein; 1465 AA.
XX AC ADN22447;
XX AC ADN22447;
XX 02-DEC-2004 (first entry)
XX Bacterial polypeptide #5100.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polypeptide.
XX

OS Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P..
XX (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 5100; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition. Improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX Sequence 1465 AA;

Query Match 5.1%; Score 163.5; DB 8; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.00048;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;
QY 2 ADAPILFGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKQK 61
DB 799 ADTVIIFSDWNPDNDLQMSRAHRIGQTKVNIYRLVTKGSVEE-----EI 845
QY 62 VENLNRLTHALLMWSGASYLFDKLDH-----FHSSETPDGSGVSPFQSDGIMDGV 108
DB 846 VERAKRKL-----LDHLVIQRMDDTGTVLSKNATASGSPFDKQLSAI 891
QY 109 IHEFPSSI--LSSKGGGENEV-----KCLLLEKHAQGYSSDSTLFGEDHIKLSDEES 160
DB 892 L-KFGAVELFKKEGEQEPEDVIDRILMGAETREAEEMKENELLSKPYANFAIDEE 950
QY 161 PNIF-----WSKLLGGKPKMVKYPSDTPQNRKR-----Q 191
DB 951 KDIAAATDEWAAII-----PEEDNRILIEERMKELAEMLAPRQKPIQ 997

cytostatic.

Homo sapiens.

WO200298899-A2.

12-DEC-2002.

03-JUN-2002; 2002WO-US017466.

05-JUN-2001; 2001US-0296076P.

10-OCT-2001; 2001US-0328605P.

22-OCT-2001; 2001US-0338733P.

15-FEB-2002; 2002US-0357253P.

15-FEB-2002; 2002US-0357600P.

(EXEL-) EXELIXIS INC.

Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP; Lioubin MN;

WPI; 2003-156840/15.

Identifying a candidate p53 pathway-modulating agent as therapeutic targets for disorders related to defective p53 function e.g. cancer by contacting an assay system having purified CHD polypeptide or nucleic acid, with a test agent.

Claim 13; Page 185-192; 278pp; English.

The present invention relates to a method for identifying candidate p53 pathway modulating agents. The method involves contacting an assay system comprising purified chromatin organisation modifier (chromo) domain helicase DNA binding proteins (CHD), nucleic acids, their functionally active fragments or derivatives, with a test agent under conditions where, but for the presence of the test agent, the system provides a reference activity. The methods are useful for identifying modulators of the p53 pathway as therapeutic targets for disorders associated with defective p53 function, such as angiogenic disorders, apoptotic disorders or cell proliferative disorders, e.g. cancer. The modulators are useful as research reagents, diagnostics and therapeutics. The invention is also useful in gene therapy. The present sequence is human chromodomain helicase DNA binding protein

Sequence 1739 AA;

Query Match 5.1%; Score 162.5; DB 6; Length 1739;

Best Local Similarity 18.9%; Pred. No. 0.00077;

Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

QY 2 ADATILFGSLNPSHDVHVEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49

DB 879 ATTVIFDSDFNDPNDLQARAHRIQKQKQVNIYLVTKGTVEEILIERAKKQVLDHL 938

QY 50 ILARQNKQKAVEN-----LNRSILTHALLMMGASLYFDKLDPHFSSTPPSGVSF 100

DB 939 VIQRMDTTKRTILENNSGRSNPNFKNEELTAILKFGAEDLPKELE-----GEES 988

QY 101 EQSINDGVTHFSSILLSSKGGSENEVKLLLEAKHAQCTYSDDSTLFGEDHIKLSDEES 160

DB 989 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQPKVANFATMEDEBEL-----EER 1038

QY 161 PNIFFSKLGGKPNMKYPSTPQNRKRVQVFE-----GSEASPKTG 203

DB 1039 PHKWDDEIL-----PEQRKKEVEERQKELBEIYMLPRISSTYKKAQTND 1085

QY 204 DGNNAKRRKASDDVTPDPRVTPPVDVDDERKASGKDHMGALESPVKVITLQSSCKSSGTG 263

DB 1086 DSDTESKQAKQSSASESED-----SDDDK-----PKRGRPRSVRKDLVEG 1130

QY 264 TLQGN-----DAFLYSMSGSHSIGIPEDML--ASQDWKIPDESQRRLHTVLKPKWAKLC 316

DB 1131 FTDAETRRFIKAYKFP-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174

QY 317 QVLHLSDACTSMVGNFLEYVNIENHRIYEBPATTFFAQIALSWIAALLVKQLSHKESLV 376

DB 1175 ELIHH--NSCVSAMQEEYEQLENASEGKGFGR-RGPTTIKISGV-QVNVKSIHQHEEPE 1230

QY 377 RANSEL-----APKCSRVEVDYIYSILSCMSKSLFLEHTQGLQFPCFGTNSK 422

DB 1231 MLHKSIPVDPEKKKYCLTCRVKAAHFDVWEGVEDDSRLLLGLIYEHGYG-NWELIKTDPE 1289

QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMNSSE-----DEECMTEKR 464

DB 1290 -----LKLTDKILPVET--DKPQKQLQTRADYLLKLARKGLEKKGAVTGGSEAKLKR 1342

QY 465 -----CSHYSTATRDIEKTIISGIKKYKKQVQKLVQEHHEKK 501

DB 1343 KPRVKKENKVPRLKBEHGIELSPPRSDNPSBEGEVKDDGLEKSPMKKKQK-KKENKENK 1401

QY 502 MELLNMYADK-----KQKLETSKVEA-----AVIRITCSTSTQVGD--- 539

DB 1402 EKQMSRKDKGDKERKKSKDKKEPKSGDAKSSSKSKRSQGPVHITAGSEPVIGDED 1461

QY 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564

DB 1462 DDLDOETFSICKERMPVKALKQDKPKGLNVQEQLEHTRNCLIKIGDRIAECLKAYS 1521

QY 565 -----LEQMHEVAKKLADEACWINRIKSWAAKVKCV 598

DB 1522 DQEHIKLWRRNLWIFVSKTFEPDARKLHKLYKMAHKRSQEE-----BQKKDDV 1572

QY 599 PTQSGNNKHF-----SGSSN-----ISQN 617

DB 1573 ---TGKKFPRPEASGSRDSLISQS 1595

RESULT 6

ID AAE36106 standard; protein; 1739 AA.

AC AAE36106;

XX 26-JUN-2003 (first entry)

XX Human chromodomain helicase DNA binding protein (CHD) #4.

XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;

KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;

KW cell proliferative disorder; chromatin organisation modifier domain;

XX cytostatic.

OS Homo sapiens.

XX Key Location/Qualifiers

EH Domain 281..327

FT Domain /note= "Chromodomain"

FT Domain 375..433

FT Domain /note= "Chromodomain"

FT Domain 487..768

FT Domain /note= "SNF2N domain"

FT Domain 831..905

FT /note= "Conserved C-terminal domain"

XX WO200298899-A2.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017466.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 22-OCT-2001; 2001US-0338733P.

XX 15-FEB-2002; 2002US-0357253P.

XX 15-FEB-2002; 2002US-0357600P.

XX

PA	(EXEL-) EXELIXIS INC.	
XX	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;	
PI	Lioubin MN;	
XX		
DR	WPI; 2003-156840/15.	
XX		
PT	Identifying a candidate p53 pathway-modulating agent as therapeutic	
PT	targets for disorders related to defective p53 function e.g. cancer by	
PT	contacting an assay system having purified CHD polypeptide or nucleic	
PT	acid, with a test agent.	
XX		
XX	Claim 13; Page 192-199; 278pp; English.	
PS		
XX	The present invention relates to a method for identifying candidate p53	
CC	pathway modulating agents. The method involves contacting an assay system	
CC	comprising purified chromatin organisation modifier (chromo) domain	
CC	helixase DNA binding proteins (CHD), nucleic acids, their functionally	
CC	active fragments or derivatives, with a test agent under conditions	
CC	where, but for the presence of the test agent, the system provides a	
CC	reference activity. The methods are useful for identifying modulators of	
CC	the p53 pathway as therapeutic targets for disorders associated with	
CC	defective p53 function, such as angiogenic disorders, apoptotic disorders	
CC	or cell proliferative disorders, e.g. cancer. The modulators are useful	
CC	as research reagents, diagnostics and therapeutics. The invention is also	
CC	useful in gene therapy. The present sequence is human chromodomain	
CC	helixase DNA binding protein	
XX		
SQ	Sequence 1739 AA;	
	Query Match	
	Best Local Similarity 5.1%; Score 162.5; DB 6; Length 1739;	
	Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;	
QY	2 ADATILFGSSLNPSHDVHKVEKIKIESCSERTKIPLRYSVCVVEKA-----L 49	
DB	879 ADTVVFSDWNPQDLQQAARHIGQKKQVNIYRLVTGVEBEIERAKKGMVLHDL 938	
QY	50 ILARQNKQKNAVEN-----LNRLTHALLMWGASYLPDKLDHFSSTPDSGVSF 100	
DB	939 VIQRMDTGTRTILENNSGRSNPNFKBELTAILFGAEDLPKELE-----GEES 988	
QY	101 EQSIMGVTHFPSSILSSKGGSENEVKLLCLLAKHAQTYSSDSTLFGEDHILKLSDEES 160	
DB	989 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQPKVANFATMEDEBEL-----EER 1038	
QY	161 PNIWFSKLLGGKNPMWKPSPDTPQNRKRVQVFE-----GSEASPTKG 203	
DB	1039 PHKOWDEII-----PEBQRKVEEERQKELEEIYMLPRINSSYTKKAQTND 1085	
QY	204 DGGNAKXKKAASDDVTDPRTVPVDDDERKASGKDHMGALSPKVITILQSSCKSSGTDG 263	
DB	1086 DSDTESKQAQRSSASESED-----SDDDK-----PKRGRPRSVRKDLVEG 1130	
QY	264 TLQGN-----DAFGLYSMGSHISGIPEDML--ASQDWGKIPESQRRLLHTVLKPKWAKLC 316	
DB	1131 FTDAIRIRFIKAYKXF-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174	
QY	317 QVILHLSDACTSMVGNFLEVVIENHRIYEPAITFQAFQIALSWIAALLVKQILSHKESIV 376	
DB	1175 ELIHF--NSCVSAMQBYEOLKENASGEGKPGKR-RGPTIKISGV-QVNVKSIHQHEEPE 1230	
QY	377 RANSEL-----AFKCSRVEVDIYISILSCMKSLFLEHTQLOQDFGTNSK 422	
DB	1231 MLHKIIPVDPEEKYKCLTCRVKAAHFDVEMGVEDDSRLLLGIYEHGYG-NWELIKTDPE 1289	
QY	423 QSVSTKLVNESLSGATVDEKINTKSMRNSSE-----DEECWTEKR 464	
DB	1290 -----LKLTDKILPVET--DKPQCGKQLQTRADYLLKLLRGLKKGKGAVTGGEAKLKKR 1342	
QY	465 -----CSHYSTATDIETISGIIKKYKQVQKLVQHEEKK 501	
DB	1343 KPRVKKENKPKLKEEHGIELSSPRHSDNPSEGEVKGDLKSPMKKKQK-KKENKENK 1401	
QY	502 MELLNMYADK-----KQKLETSKSVFA-----AVIRITCSRTSTQVGD--- 539	
DB	1402 EQOMSSRXDKGDKERKKSQKKEKPKSGDAKSSSKSRSQGPVHTAGSEVPVIGEDD 1461	
QY	540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564	
DB	1462 DDLDOETFSICKERMVRPVKKALKQLDKPKDGLNVOEQLEHTRNCLLKIGDRIAECLKAYS 1521	
QY	565 -----LEQMEHVAKKLAEDACWINRIKSWAAKLKVCV 598	
DB	1522 DOEHIKLWRNLWIFVSKTFEDARKLHKLTKMAHKKRSQEEE-----EQKKDDV 1572	
QY	599 PIQSGNNKHF-----SGSSN-----ISQN 617	
DB	1573 ---TGGKKPFRPEASGSRDSLSISQS 1595	
RESULT 7		
ID	ADK60208	
ID	ADK60208 standard; protein; 1739 AA.	
XX	ADK60208;	
AC		
XX	06-MAY-2004 (first entry)	
DT		
XX	Angiogenesis differentially expressed protein GS-P32.	
DE		
XX	vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;	
KW	antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;	
KW	angiogenesis; endothelial cell; diagnosis; tumor vascularization;	
KW	retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;	
KW	ovary hyperstimulation; psoriasis; endometriosis; restenosis;	
KW	angioplasty; cicatrization; peripheral vascular disease; hypertension;	
KW	vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;	
KW	ischemia; angina; myocardial infarction; chronic heart disease;	
KW	cardiac congestion; macular degeneration; osteoporosis.	
XX		
OS	Homo sapiens.	
XX		
PN	FR2836687-Al.	
XX		
PD	05-SEP-2003.	
XX		
PF	11-APR-2002; 2002FR-00004546.	
XX		
PR	04-MAR-2002; 2002FR-00002717.	
XX		
PA	(GENE-) GENE SIGNAL.	
PA	(ALMA/) AL MAHMOOD S.	
PI	Colin S, Schneider C, Al Mahmood S;	
XX		
XX	WPI: 2004-013912/02.	
DR	N-PSDB; ADK60458.	
XX		
PT	Compositions for diagnosing, prognosing and treating angiogenic disorders	
PT	including tumor vascularization and heart disease, comprise nucleic acid	
XX	or polypeptide differentially expressed in angiogenesis.	
XX		
PS	Claim 7; SEQ ID NO 84; 424pp; French.	
XX		
CC	The invention relates to a novel pharmaceutical composition active on	
CC	angiogenesis comprising an endothelial cell nucleic acid whose expression	
CC	is induced by an angiogenic factor and inhibited by an angiostatic agent	
CC	or its complement or fragment, a polypeptide sequence encoded by the	
CC	nucleic acid or its fragment, a molecule capable of inhibiting expression	
CC	of the nucleic acid or a molecule which binds to the polypeptide	
CC	sequence. The invention is used to diagnose, prognose or treat an	
CC	angiogenic disorder in a mammal, particularly a human. The disorder is	
CC	particularly tumor vascularization, a retinopathy, rheumatoid arthritis,	
CC	Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,	
CC	endometriosis associated with neovascularization, restenosis due to	

[illegible]

RESULT 11	
ADN20557	
ID	ADN20557 standard; protein; 1015 AA.
XX	
AC	ADN20557;
XX	
DT	02-DEC-2004 (first entry)
XX	
DE	Bacterial polypeptide #3210.

microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Query Match 5.0%; Score 161; DB 8; Length 1015;
Best Local Similarity 20.1%; Pred. No. 0.00045;

```

RESULT 12
AAU14615
ID AAU14615 standard; protein; 954 AA.
XX AC
XX AAU14615;
XX 24-OCT-2001 (first entry)
XX DE
XX Novel bone marrow polypeptide #14.
XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX OS
XX Homo sapiens.
XX WO200157187-A2.
XX PN
XX 09-AUG-2001.
XX PD
XX 05-FEB-2001; 2001WO-US0003782.
XX PF
XX 03-FEB-2000; 2000US-00496914.
XX PR
XX 20-JUN-2000; 2000US-00598075.
XX PR
XX 19-JUL-2000; 2000US-00620325.
XX PR
XX 30-NOV-2000; 2000US-0250683P.
XX PA
XX (HYSE-) HYSEQ INC.
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX WPI; 2001-488875/53.
XX DR
XX N-PSDB; AAS22920.
XX DR
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
XX gene therapy.
XX Claim 10; Page 215-217; 392pp; English.
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
XX invention. The proteins and corresponding coding sequences may be used in
XX the prevention, diagnosis and treatment of diseases associated with
XX inappropriate bone marrow polypeptide expression. For example, to treat
XX disorders associated with decreased expression by rectifying mutations or
XX deletions in a patient's genome that affect the activity of the
XX polypeptides by expressing inactive proteins or to supplement the
XX patient's own production of the polypeptides. Additionally, the nucleic
XX acids may be used to produce the polypeptides, by inserting the nucleic
XX acids into a host cell and culturing the cell to express the protein. The
XX nucleic acid and its complementary sequences may also be used as DNA
XX probes in diagnostic assays to detect and quantitate the presence of
XX similar nucleic acid sequences in samples, and therefore which patients
XX may be in need of restorative therapy. The proteins may also be used as
XX antigens in the production of antibodies against bone marrow proteins and
XX in assays to identify modulators of their expression and activity. The
XX anti-bone marrow protein antibodies and antagonists may also be used to
XX down regulate expression and activity. The antibodies may also be used as
XX diagnostic agents for detecting the presence of the protein in samples
XX (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins may be
XX used to regulate haematopoiesis activity, and consequently in the
XX treatment of myeloid or lymph cell disorders; in tissue regeneration,
XX such as wound healing; as a nutritional supplement; and in treatment of
XX immune disorders such as severe combined immunodeficiency (SCID)
XX SQ
Sequence 954 AA;
Query Match 4.9%; Score 156; DB 4; Length 954;
Best Local Similarity 18.9%; Pred. No. 0.0011;
Matches 131; Conservative 105; Mismatches 233; Indels 224; Gaps 28;
XX 93 TPDSGVSEFQSGIMDGVITHEFS-----SILSSKGGEEVVKLCLLLEAKHAQGTYSDDSTLF 148

```

```

19 SPGAGSSLSAVGKG--ROGSGARPGVPKEEENEKPEKTCRNMVAPKADPT----- 68
149 GEDHIKLSDEESPNIWFWSKLLGG-----KNPMWKY-----PSDTPQ-----RNR 187
69 GSEPAKASEKEAPE---DITDAGEGSRREAAKEPKKASALEEGSSDASQLEISEHM 125
188 KRVOYFEGSEASPKTGDGNNAKRRKKASDDVTDPRTVDP-----PVDDDERK 234
126 KEPQLSDSIASDPKSPHGLDFGFRSRISEHLLDVLSPVLGGACRQAQPLGIEDKDD 185
235 ASGKDHMGALLES-----PKVITLQSSCKS-----SGTDGTLT----- 266
186 QSSQDELQSKQSGLEERYHRLSPPLPHEERAQSPRSLATEBPFGPEGQEWKEABE 245
267 -GNDAFGLYSM-----GSHISGIPEDMLA----- 289
246 LGEDSAASLSLQSLQREQAPSPAAACEKGEQHSQAELGPGQEADEEEKVAVSPTP 305
290 -----SDWGKIPDESQRRLHTVLPKPM-----AKLQCO-- 317
306 PVSPEVSTPEVPAPPEQLSEAAKAMEEAVAQVLEQDQRHLLSEKQEKMQQLREKLQEE 365
318 ---VLHL---SDACTSMVGNFLEYVIEHRIYEEPTTFOAFOIALSWIAALLVKQLSH 371
366 EEEILRHQOQESLSLRERLQKAYE-----EEEARMREESQRLSWLRA-QVQSSTQA 419
372 KESLVRANSELPKCSRVEVDYIYLSQWK-----SLFLEHTQGL-----QPCFGTNSK 422
420 DEDQIIRAEQASLQKUREELE-----SQKAEASLEQKNQMLQOLKEEIEA-SEKSE 472
423 QSVVST-----KLVNESLSGATVRDEKINTKSMRNSDEECMTKRCRSHYSTATRDI- 475
473 QAAALNAAKEKALQQLREQLEGE--RKEAVATLEKHSAB-----LERLCSLEAKHREV 525
476 ----EKTIISGIGKKYKQVKQVQKLVQEHKEK-----KMLLNMYADKKQKLE----- 516
526 SSLQKKIQBAQOQKEEAQQLQCLGQVHRVHQSYHVAGYEHLSLLREKQVEGEHER 585
517 -----TSKSVEAAVIRITCSRTSTQVGDLLKLDHNYERKPFDEIKSEKNEC 561
586 RLDKMKEEHQVWAKAREQYEAERKQRAELHLGTGELERLQRAHERLETVRQSHKR 645
562 LKSLQEQHVEVAKKKLAEDAEACWINRIKSWAAKL 594
646 LEDLRERHREQERKQLDLELLETRAKDKVKARL 678
RESULT 13
ABR61599
ID ABR61599 standard; protein; 2228 AA.
XX AC
XX ABR61599;
XX 15-JAN-2004 (first entry)
XX Human golgin-245 splice variant 1 polypeptide.
XX Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
XX antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX Homo sapiens.
XX WO2003087403-A2.
XX 23-OCT-2003.
XX 16-APR-2003; 2003WO-EP003958.
XX 16-APR-2002; 2002EP-00008553.
XX 16-APR-2002; 2002US-0372424P.
XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.

```

XX Von Der Kammer H, Pohlner J;
XX WPI; 2003-945345/78.
XX N-PSDB; ACF58150.
XX Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
XX
XX Claim 18; Fig 5; Opp; English.
XX
XX The invention relates to diagnosing or prognosticating a
CC neurodegenerative disease in a subject, or determining whether a subject
CC is at increased risk of developing the disease. The method involves
CC determining a level and/or an activity of a transcription or translation
CC product of a gene coding for golgin-245 and/or a fragment, derivative or
CC variant in a sample from the subject and comparing the level and/or the
CC activity to a reference value. The composition and methods are useful in
CC diagnosing, preventing and/or treating neurodegenerative diseases, such
CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
CC useful for screening, testing and validating compounds, agents and
CC modulators in the development of diagnostics and therapeutics to treat
CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
CC antibodies are used for detecting the pathological state of a cell in a
CC sample from a subject. The present sequence represents a human golgin-245
CC splice variant 1 polypeptide
XX
XX Sequence 2228 AA;
SQ
Query Match 4.8%; Score 154.5; DB 7; Length 2228;
Best Local Similarity 19.7%; Pred. No. 0.0055;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;
QY 13 NFSHDVHVEKI-KTESCE-----RTKIFRLYSVCTVEKALILARONKRONKAVEN 64
DB 1552 NQKOIEHKELVQKLFHGFELGEEKDNRYK-----BAEEKILTLENQVYSMKAELET 1603
QY 65 LNRSLSLTHALLMGA-----SYLFDKLDHFSSETPDGVSFEQIMDGVITHFSSLTSSK 119
DB 1604 KKELEHVNLSVKSKEEELKALDRLSESAAKLAELKRAKQKIA-AIKKQLLSQWEEK 1662
QY 120 GGEENEVKLCLEAKHAQGYTSSDTLFGEDHIKLSD-EESPNI FWSKLLGGKPNMVKY 178
DB 1663 -----EEQYKKGTSLSSEL---NTKLQEREREVHILLEELKLSVESQ--- 1702
QY 179 PSDT---PORNKRVOYFEGSEASPKTGGGNAKRRKASDDVTPRVTDPV-----D 229
DB 1703 -SETLIVPRSAKNVAAYTQEESDSQ---GCYQKTYEEKISVLQRNLTEKELQLRVGQ 1757
QY 230 DDERKASGDHMGALSPKVTITLQSSCKSGTDTLDG-----NDAPGLYMGSHI 280
DB 1758 EKEETVSHFEMRCQVQERELKLEHAKQHEQDSMIGHLQEELEKKNKYSILI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLCQVLHLS-----DACTSMVGNFLE 334
DB 1817 EREGGKNNTQAQNLENVDDVQKTLQ-----EKELTCCILEQKIKELDSC-----LV 1864
QY 335 YVIENHRI-YEPATTFOAFQALSWI-----AALLVQILSHKES 374
DB 1865 ROKEVHRVMEELTSKYELQ-ALQOMGRNKPTEELLEENTEEKSHLVQPKLLSNMEA 1923
QY 375 LVANSELAFKCSRVEVD-----YIYSILSCWKSFLFLEHTQLOPDCFGTNSQSWST 428
DB 1924 ----QNDLEFKLAGAREKQKLGKELVIRLOKLMRLKKEHQELE-----ILKK 1969
QY 429 KLVNLSLGATVDEKINTK-----SMRNSSEDEECMTEKRCSHYSTPATRIDIKTSIGIKK 484
DB 1970 EYDQE-----REEKIQEQEBDELKHNSTLKLQLMREFN-TQLAQKEQELETMTKETIN 2021
QY 485 KYTKQVQKLVQEHKEKMLLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDKL--- 542

DB 2022 KAEVEAEELLESQBETNQLLKKAIEKODDLKRTAKRYEEIILDAREEENTAKVRLQTL 2081
QY 543 --LDHNYERKFEIKSEKNECLKSLEQMEHVAKK-----KLADEACWINRIKSWAAK 593
DB 2082 EELQKKYQKLEQENPGNDNVITMELQTLQAOKTLLSDSKLKEQE--FRSQIHNLDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
DB 2140 LKKYEKNVYATTGTPYKGNLYH 2163
RESULT 14
ABU07445
ID ABU07445 standard; protein; 2230 AA.
AC ABU07445;
XX 28-JAN-2003 (first entry)
DT Protein differentially regulated in prostate cancer #48.
DE Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
OS Homo sapiens.
XX WO200281638-A2.
XX 17-OCT-2002.
XX 08-APR-2002; 2002WO-USO10824.
XX 06-APR-2001; 2001US-0281731P.
XX 06-APR-2001; 2001US-0281732P.
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Jay G;
XX WPI; 2003-058520/05.
XX N-PSDB; ABX10347.
XX Novel genes which are differentially regulated in prostate cancer, useful
PT for diagnosing prostate cancer in prostate tissue sample and assessing
PT therapeutic or preventive intervention in prostate cancer patients.
XX Claim 1; Page 299-306; 416pp; English.
XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) Is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) Is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity. (I)
CC is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type

CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (1) can be used as target for therapy or drug
CC discovery. (1) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (1) is useful
CC in therapeutic applications to treat prostate cancer. The identification
CC of specific genes, and groups of genes, expressed in pathways
CC physiologically relevant to prostate cancer permits the definition of
CC functional and disease pathways and the delineation of targets in these
CC pathways which are useful in diagnostic, therapeutic, and clinical
CC applications. This is the amino acid sequence of a protein differentially
CC regulated in prostate cancer
XX
SQ Sequence 2230 AA;

Query Match 4.8%; Score 154.5; DB 6; Length 2230;
Best Local Similarity 19.7%; Pred. No. 0.0055;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;

QY 13 NPSHDVKHVEKI-KIESCSSE-----RTKIFRLYSVCTVEEKALILARQKQKQKAVEN 64
DB 1552 NQOKDIEHKELVQKLFQELGEEKDNVRK-----EAEKILTLENQVYSMAKAELET 1603
QY 65 LNRSILTHALLMGA-----SYLFDKLDHFHSSSTPDGVSFQSGIMDGVTHIEFSSILSSK 119
DB 1604 KKELEHVNLSVKSKEELKALEDRLSESAKLAELKRAEQKIA-AIKKQLLSQMEEK 1662
QY 120 GGBENEVKLCLEAKHAQGYSSDSTLFGEDHIKLSD-EESNPIFWSKLLGKGNPMWKY 178
DB 1663 -----EEQYKKGTSHELSEL-----NTKQEREREVHILLEEKLSVSESSQ--- 1702
QY 179 PSDT---PQNRKRQVYFEGSEASPKTGDGNAKRAKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPSAKNVAAYTEQEEADSQ---GCQVKTVEEKISVLQRNLTEKEKLLQRVGQ 1757
QY 230 DDERKASGDKHGALESPKVTITLQSSCKSGTGTLOG-----NDAFLYSGMSHI 280
DB 1758 EKETVSSHPEMRCQVQERIKLEHAEAKQHEQDSMIGHLOELEEKNKYSILI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLCQVLHLS---DACTSMVGNFLE 334
DB 1817 EKEGKNNIQAKQNLNVFDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
QY 335 YVTENHRI-YEEPAITFOAFQIALSWI-----AALLVQKILSHKES 374
DB 1865 RQKEVHRVMEELTSKYELQ-ALQOMDGRNKPTELLEENTSEKSKSHLVQPKLLSMEA 1923
QY 375 LVPANSELAFKCSRVED-----YIYSILSCMSLPLEHTQGLQDPCFTGNSQSVST 428
DB 1924 ---OHNDLEPKLAGAREKQKGLKEIVRLQKDLRLMRKEHQQELB-----ILKK 1969
QY 429 KLVNESLSGATVRDEKINTK---SMRNSSEDECMTEKRSCHYSTATRDIEKTISGIKK 484
DB 1970 EYDQE-----REKIKQEDLELKHNTSLQLMREFN-TOLAQKEQELEMTIKETIN 2021
QY 485 KYKQVQKLVQHEHEKQELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGDKL-- 542
DB 2022 KAQVEAEELLESQHEETNQLKKIAEKDDLKRTAKRYEILDARBEEMTAKVRDLQTLQ 2081
QY 543 --LDHNYERKFDIKSEKNECLSKLEOMHEVAK-----KLAEDCAWINIKSWAAK 593
DB 2082 BELQKYQQKLEQEPNGDNDVTIMELOTQAQKTTLISDSKLKEQE--FREQIHLEDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
DB 2140 LKYEKNVYATTGTTPYKGNLYH 2163

RESULT 15

ABR61600

ID ABR61600 standard; protein; 2230 AA.

XX

AC ABR61600;

XX

DT XX 15-JAN-2004 (first entry)
DE XX Human golgin-245 splice variant 2 polypeptide.
XX XX Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
KW XX antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX OS Homo sapiens.
XX PN WO2003087403-A2.
XX PD 23-OCT-2003.
XX PF 16-APR-2003; 2003WO-EP003958.
XX PR 16-APR-2002; 2002EP-00008553.
XX PR 16-APR-2002; 2002US-0372424P.
XX PA (EVOT-) EVOTEC NEUROSCIENCES GMBH.
XX PI Von Der Kammer H, Pohlner J;
XX PI WPI: 2003-845345/78.
XX DR N-PSDB; ACF58151.
XX DR
XX PT Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
XX PS Claim 18; Fig 7; Opp; English.
XX CC The invention relates to diagnosing or prognosticating a
CC neurodegenerative disease in a subject, or determining whether a subject
CC is at increased risk of developing the disease. The method involves
CC determining a level and/or an activity of a transcription or translation
CC product of a gene coding for golgin-245 and/or a fragment, derivative or
CC variant in a sample from the subject and comparing the level and/or the
CC activity to a reference value. The composition and methods are useful in
CC diagnosing, preventing and/or treating neurodegenerative diseases, such
CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
CC anyotrophic lateral sclerosis. The recombinant, non-human animal is
CC useful for screening, testing and validating compounds, agents and
CC modulators in the development of diagnostics and therapeutics to treat
CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
CC antibodies are used for detecting the pathological state of a cell in a
CC sample from a subject. The present sequence represents a human golgin-245
CC splice variant 2 polypeptide
XX SQ Sequence 2230 AA;

Query Match 4.8%; Score 154.5; DB 7; Length 2230;
Best Local Similarity 19.7%; Pred. No. 0.0055;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;

QY 13 NPSHDVKHVEKI-KIESCSSE-----RTKIFRLYSVCTVEEKALILARQKQKQKAVEN 64
DB 1552 NQOKDIEHKELVQKLFQELGEEKDNVRK-----EAEKILTLENQVYSMAKAELET 1603
QY 65 LNRSILTHALLMGA-----SYLFDKLDHFHSSSTPDGVSFQSGIMDGVTHIEFSSILSSK 119
DB 1604 KKELEHVNLSVKSKEELKALEDRLSESAKLAELKRAEQKIA-AIKKQLLSQMEEK 1662
QY 120 GGBENEVKLCLEAKHAQGYSSDSTLFGEDHIKLSD-EESNPIFWSKLLGKGNPMWKY 178
DB 1663 -----EEQYKKGTSHELSEL-----NTKQEREREVHILLEEKLSVSESSQ--- 1702
QY 179 PSDT---PQNRKRQVYFEGSEASPKTGDGNAKRAKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPSAKNVAAYTEQEEADSQ---GCQVKTVEEKISVLQRNLTEKEKLLQRVGQ 1757
QY 230 DDERKASGDKHGALESPKVTITLQSSCKSGTGTLOG-----NDAFLYSGMSHI 280

```
Db 1758 EKEETVSHFMRQVQYERLIILEHAEAKQHEQDSMIGHLQBELEBKKNKYSLI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKI PDESORRLHTVLKPKMAKLCQVLHLS-----DACTSWVGNFLE 334
Db 1817 EKEGGKNNIQAQNLNVFDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
QY 335 YVIENHRI-YESPATTFQAPQIALSWI-----AALLVQIILSHKES 374
Db 1865 ROKEVHRVEMEELTSKYELQ-ALQOMDGRNKPTELLEENTBEKSKSHLVQPKLLSNMEA 1923
QY 375 LVRANSELAFKCSRVEVD-----YIYSILSCWKSIFLEHTQGLQDFCGTNSKQSVST 428
Db 1924 ---QHNDLEFKLAGAREKQKLGKEIVRLQKDLRMURKEHQOELE-----ILKK 1969
QY 429 KLVNESLGGATVRDEKINTK---SMRNSSEDEECMTEKRCSHYSTATRDIENTISGIKK 484
Db 1970 EYDQE-----REEKIQOEDELKHNSTLKLQMRFN-TQLAQKEQELEWTIKETIN 2021
QY 485 KYKQVQKLVQHEHEKMBELLNMYADKOKLETSKSVEAAVIRITCSRTSTQVGDJL-- 542
Db 2022 KAQVEAEELSHQEBETNQLKKAIEKDDDLKRTAKRYEEILDAREEEMTAKVRDIQTOL 2081
QY 543 --LDHNYERKFEIKSEKNECLSKLEQMEHVAKK-----KLADEACWINRIKSWAAK 593
Db 2082 BELQKKYQOKLEQENPGNDNVTIMELQTLAQKTTLISDSKLKQE--FREQIHLEDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
Db 2140 LKYEKNVYATTVGTPYKGNLYH 2163
```

Search completed: July 12, 2005, 00:58:29
Job time : 67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:11:48 ; Search time 16.6667 Seconds
(without alignments)
2781.424 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILFGSLNPSHDVXH.....SGNNKHFGSSNISONAPDV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	162.5	5.1	1739	4	US-09-976-594-76
2	162.5	5.1	1739	4	US-09-538-092-824
3	162.5	5.1	1740	4	US-09-949-016-8860
4	150	4.7	1939	4	US-09-949-016-8925
5	150	4.7	1942	4	US-09-949-016-8135
6	147	4.6	2733	4	US-09-949-016-11433
7	146.5	4.6	630	4	US-09-248-796A-20275
8	145.5	4.5	1648	4	US-09-248-796A-14286
9	143	4.4	1960	4	US-09-538-092-1077
10	143	4.4	1960	4	US-09-949-016-10872
11	142.5	4.4	1427	4	US-09-538-092-1044
12	140.5	4.4	1972	4	US-08-875-435B-4
13	140	4.4	3259	4	US-09-949-016-6507
14	138	4.3	1836	4	US-09-949-016-7432
15	136.5	4.2	3210	4	US-09-538-092-1154
16	136.5	4.2	3248	1	US-08-353-700-1
17	136.5	4.2	3248	5	PCT-US95-16216-1
18	136.5	4.2	8991	4	US-08-714-741-32
19	135.5	4.2	2482	1	US-08-328-254-6
20	135	4.2	1939	3	US-09-310-187A-1
21	135	4.2	1939	4	US-09-538-092-917
22	134	4.2	1388	4	US-09-976-594-296
23	134	4.2	1695	4	US-09-866-108A-15753
24	133.5	4.2	2186	4	US-09-949-016-10828
25	133.5	4.2	2349	4	US-09-538-092-914
26	133	4.1	901	4	US-09-248-796A-14747
27	133	4.1	1388	2	US-08-685-576-1

ALIGNMENTS

RESULT 1
US-09-976-594-76
; Sequence 76, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1
US-09-976-594-76

Query Match	5.1%;	Score 162.5;	DB 4;	Length 1739;
Best Local Similarity	18.9%;	Pred. No. 2.1e-05;		
Matches	152;	Conservative 116;	Mismatches 259;	Indels 279; Gaps 35;
QY	2	ADAFILFGSLNPSHDVXVEKIKIESCSERTKIFRLYSVCTVEKA-----L 49		
Db	879	ADTVIFDSDWNPQDLQARAHRIQKQVNIYRLVTGTVVEEIIERAKKXVLDHL 938		
QY	50	ILARONKRONKAVEN-----LNRSUTHALLMMGASYLPDKLDHPSHSETPDGVSF 100		
Db	939	VIQRMDDTGTILENNSGRNSNPFNKEELTAILKGAEDLFKELE-----GEES 988		
QY	101	EQSIMGVTHFESSILSSKGGGENEYKLCLELAKAOGTYSSDSTLFGEDHIKLSDEES 160		
Db	989	EPOEMD-----IDEILRLAETRENEVSTSATDELL--SQPKVANFATMEDEEEL-----EER 1038		
QY	161	PNIFWSKLGCKNPMWKPSTDPORNRKVQVPE-----GSEASPKTG 203		
Db	1039	PHKWDDEII-----PEQRKVVBEERKELEEIYMLPRISSTTKAQTND 1085		
QY	204	DGNNAKRRKKASDDVTPDPRVTPDPPVDDERKASGDHMGALSPKVITLQSSCKSGTDG 263		
Db	1086	DSPTESKRAQRSASESETED---SDDDKK-----PKRRGRPRSVRKDLVEG 1130		
QY	264	TLDGN-----DAPGLYMSGHISGIPEDML--ASQWQKIPDSQRRLTLVLPKMAKLC 316		

```
Db 1131 FTDAEIRRFKAYKFP-----GLPLERLECIARD-AELVDKSVADL-----KRLG 1174
QY 317 QVVLHLSDACTSMVGNFLEYVNIENHRIYEBPATTFFOAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQYEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIIOHEEPE 1230
QY 377 RANSEL-----AFKCSRVEDYIYISILSCMKSLFLEHTQGLQDFCFGTNSK 422
Db 1231 MLHKSIPVDPPEKKYKCLTRVKAHFVDEWGVDDSRLLGLIYEHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLINESLSGATVDEKINTKSMNSSE-----DEECMTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGEAKLKR 1342
QY 465 -----CSHYSTATRDIEKTIISGIIKKYKQVQKLVQEHHEKK 501
Db 1343 KPRVKENKVPRLKEHGIELSPRHSNPNSEGEVDDGLKSLPKMKKQK-KKENK 1401
QY 502 MELLNMYADK-----KQLETSKSVEA-----AVIRITCSRTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKOKKEPKSGDAKSSKRSQGPVHITAGSEPVPIGEDD 1461
QY 540 -----IKLIDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQDQFTSICKERMVRPVKALKQDKPKGLNVQGLEHTNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMEHVAKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DOEHIKLRNRLWIFVSKTFEPDARKLHLKYMAHKRSQEBE-----EQKKODV 1572
QY 599 PIQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGKKPFRPEASGSRDLSISQS 1595

RESULT 2
US-09-538-092-824
; Sequence 824, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 824
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number O14647
US-09-538-092-824

Query Match 5.1%; Score 162.5; DB 4; Length 1739;
Best Local Similarity 18.9%; Pred. No. 2.1e-05;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

QY 2 ADAPILFGSLNPSHDVKEVKEKIESCSERTKIFPLYSVCTVEEKA-----L 49
Db 879 ADTVVIFSDWNPQNDLQAARHRTGQKKQVNIYRLVTKGTVEESIIERAKKMKWLDHL 938
QY 50 ILARONKRONKAVEN-----LNRLTHALLMWGASVLPDKLDHFSHSPDPSGVGF 100
Db 939 VIQRMDTTGRTILNNSGRSNPNFKNBELTAILKFCAEDLFKELB-----GEES 988
```

```
QY 101 EQSIMGDVIHEPSSILSKSGEENEVYKLCLLLEAKHAQGTYSDDSTLFGEDHILSDERS 160
Db 989 EFQEND-----IDEILRLAETRENEVSTSATDELL-SQFKVANFATMEDEEL---EBR 1038
QY 161 PNIWFSKLLGGKNPMWKYPSDTPQNRKRVQYFE-----GSEASPKTG 203
Db 1039 PHKWDDEII-----PEQRKKVSEERQKELEETIYMLPRINSSTKKAQTND 1085
QY 204 DCGNAKKRKKASDDVTPRVDDDERKASGDHMGALSPKVITLQSSCKSGTGD 263
Db 1086 DSDTESKQAQRSSASETED---SDDDK-----PKRGRPRVRKDLVEG 1130
QY 264 TLIDGN-----DAFGLYSMSGSHISGIPEDWL--ASQDWGKIPDESORRLTLVKPKMAKLC 316
Db 1131 FTDAEIRRFKAYKFP-----GLPLERLECIARD-AELVDKSVADL-----KRLG 1174
QY 317 QVVLHLSDACTSMVGNFLEYVNIENHRIYEBPATTFFOAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAMQYEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIIOHEEPE 1230
QY 377 RANSEL-----AFKCSRVEDYIYISILSCMKSLFLEHTQGLQDFCFGTNSK 422
Db 1231 MLHKSIPVDPPEKKYKCLTRVKAHFVDEWGVDDSRLLGLIYEHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLINESLSGATVDEKINTKSMNSSE-----DEECMTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGEAKLKR 1342
QY 465 -----CSHYSTATRDIEKTIISGIIKKYKQVQKLVQEHHEKK 501
Db 1343 KPRVKENKVPRLKEHGIELSPRHSNPNSEGEVDDGLKSLPKMKKQK-KKENK 1401
QY 502 MELLNMYADK-----KQLETSKSVEA-----AVIRITCSRTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKOKKEPKSGDAKSSKRSQGPVHITAGSEPVPIGEDD 1461
QY 540 -----IKLIDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQDQFTSICKERMVRPVKALKQDKPKGLNVQGLEHTNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMEHVAKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DOEHIKLRNRLWIFVSKTFEPDARKLHLKYMAHKRSQEBE-----EQKKODV 1572
QY 599 PIQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGKKPFRPEASGSRDLSISQS 1595
```

RESULT 3

```
US-09-949-016-8860
; Sequence 8860, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8860
; LENGTH: 1740
; TYPE: PRT
; ORGANISM: Human
```

US-09-949-016-8860

Query Match 5.1%; Score 162.5; DB 4; Length 1740;
Best Local Similarity 18.9%; Pred. No. 2.1e-05;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

QY 2 ADAPILGSSLNPSHDVKKVEKIKIESCSRTKIFRLYSVCTVEKA-----L 49
DB 880 ADTWTFDSWNPNOLQARAHRIQKQKQVNIYRLVTKGTVEEIERAKKQWLDHL 939
QY 50 ILARONKRONKAVEN-----LNRLTHALLMGCASYLFDKLHFSHSETPDGVSF 100
DB 940 VIQEMDTTGTEILNENSGRNSPNFKKEELTAILKFGAEDLFKELE-----GEES 989
QY 101 EQSIMDGVIFHSILSSKGEENEVKLCILLLEAKHAQGTYSSTDLFGSDHILKSDSES 160
DB 990 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQFVANFATMEDEEL-----EER 1039
QY 161 PNIFWSKLLGGKPMKYPSTQNRKRVOYEE-----GSEASPKTG 203
DB 1040 PHKDWDEII-----PEQRKKVVEEERQKLEBEIYMLPRISSSTKKAQTNDS 1086
QY 204 DGNNAKRRKASDDVTPDPRVTPDDDERKASGDHMGALBSPKVTITLQSSCKSGTGD 263
DB 1087 DSSTESKQQRSSASETED---SDDDKK-----PKRGRPRSVRKDLVEG 1131
QY 264 TLDGN-----DAFLYSMGSHIGIPEDML--ASQDWGKIPDSORRLHTVLKPKMAKLC 316
DB 1132 FTDAETRRFTKAYKF-----GLPLERLECIARD-AELVDKSVADL-----KRLG 1175
QY 317 QVHLSDACTSMGVNFLEYVNIHRIYEEPAITFOAFQIALSWIALLVKQILSHRESLV 376
DB 1176 ELTH--NSCVSAMQRYEQLKENASEGKPGKR-RGPTTIKISGV-QVNVKSIHQHEEPE 1231
QY 377 RANSEL-----AFKSRVEVDYIYVLSLSCMKSLEHTQGLQPCFGTNSK 422
DB 1232 MLHKSIPVDPPEKKYCLICRVKAAHFDVWGVGVEDDSRLLLGIYEHYG--NWELIKTDPE 1290
QY 423 QSVSTKLVNESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
DB 1291-----LKLTDKILPVET--DKKPGQLQTRADYLLKLLKGLKKGAVTGGEEAKLKKR 1343
QY 465-----CSHYSTATRDIETKISGIIKKYKQVQKLVQHEBEKK 501
DB 1344 KPRVKKENKVPRLKEEGHIELSSPRSDNPSEGEVKDGLGKSPMKKQK-KKENKENK 1402
QY 502 MELLNMYADK-----KQLETSKSVRA-----AVIRITCSRTSTQVGD--- 539
DB 1403 EKQMSRKQEGDKERKSKDKKPKSGDAKSSKSKSGQGPVHITAGSEPVPTIGEDD 1462
QY 540-----LKLDPH-----NYERKFD-----EIKSEKNECLKS-- 564
DB 1463 DDLQETFSICKERMPVKALKQDKPKGLNVQLEHTRNCLLKIGDRIAECCLKAYS 1522
QY 565-----LEQHEVAKKKLADEACWINRIKSWAAKLVKCV 598
DB 1523 DQSHIKLWRNLWIFVSKFTEFDARKLHLYKMAHKRSQEEB-----EQKKDDV 1573
QY 599 PIQSGNKHFF---SGSSN---LSON 617
DB 1574 ---TGGKKPFRPEASGSRDLSLSQS 1596

RESULT 4

US-09-949-016-6925
; Sequence 6925, Application US/09949016
; Patent No. 6812339; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CLO01307

; CURRENT APPLICATION NUMBER: US/09/949,016

RESULT 5

US-09-949-016-8135

; Sequence 8135, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6925
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6925

Query Match 4.7%; Score 150; DB 4; Length 1939;
Best Local Similarity 21.0%; Pred. No. 0.00036;
Matches 144; Conservative 108; Mismatches 262; Indels 172; Gaps 30;

QY 22 EKIKIE-SCSERTKIFRLYSVCTVEKALILARONKRONKAVENLNSLTHALLMGCASY 80
DB 943 KKKKLEDECSLKK-----DIDDLLETLAKVEKEKH-ATENKVKNLTEE--MAGLDE 991
QY 81 LFDKLHFSHSETPDGVSFEQSIMDGVIEH--FSSILSK-----GGEENEVK 127
DB 992 TIAKL-----TYEKKALQEAHQOTLDDQMBEEDKVNTLTAKTKLEQQVDDLEGSLEQEK 1047
QY 128 LCILLLE--AKHAQG--TYSSDSTLFCG--DHIKLSDESPNI FWSKLLGGKPMKYPST 182
DB 1048 LCMDLERAKRLEGLDLKLAQESTMDTENDKQNLNEKLLKKEFEMSNLQKIEDEQALAMQ 1107
QY 183 PQNRKRKVQ-YFEGSASPKTGDGNAKRRKASD-----DVTD-----PRVTPDPVD- 229
DB 1108 LQKKIKELQARTELEEEIEAEERASRAKAKORSDLSELEISERLEEAGGATSQAIEL 1167
QY 230 DIERKASGDHMGALBSPKVTITLQSSCKSGTDTGLDNDADFGLYSMGSHISIP----- 284
DB 1168 NKKREABFOQMRDLE-----ESTLQHEATAAALRRKHADSVABLKGQIDSLQVRKQK 1220
QY 285-----EDMLASQDWGKIPDESORRLHTVL 308
DB 1221 LEKESKELQWEINDLASNMETVSKAKANFEKMCRTLEDQLSEI---KTYEEEOORLINEL 1277
QY 309 KPRMAKL-----COVLHLSDACTSMVGN-----FLEYVNIHRIYEE-----PATTFOAF 353
DB 1278 SAQKARLHTSEGEFSRQLDEKDAWVSQLSRGQAFTQOIIEELKRLQLEETKAKSTLHAHAL 1337
QY 354 QIALSWIALLVKQILSHYES-----LVANSELAFKCSRVVDYIYVLSLSCMKSFL 406
DB 1338 QSA-RHDCDLLREQVEEEOEAKAELQRGMSKANSEVAOWRTKYETDAI----- 1384
QY 407 EHTQGLQPCFGTNSKQSVVSTKLVNESLSGATVRDEKINTK-----SMRNSSEDEEC 459
DB 1385 QRTLEE-----EAKKLAQRLQDAEEHVEAVNSKASCASLEKTKQRLQNEVEDL 1432
QY 460 M-----TEKCRSHYSTARDIETKISGIIKKYKQVQKLVQHEBEKK-----MELLNMY 508
DB 1433 MIDVERSNAACIALDKQNFQKVLAEWKQVEETQAELEASQKESRSLSSTELFKVKNAV 1492
QY 509 ADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKPEDEIKSEKNECLKSLEQM 568
DB 1493 BESLOHLETLKR-ENKNLQOEISDLTEQIAEGGKHIHELEKVKQKLDHEKSEKLSQTSLEA 1551
QY 569 -----HEVAKKKLADEACWINRIKS 589
DB 1552 EASLEHEEGKILRIQLE---LNQVKS 1574

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8135
; LENGTH: 1942
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8135

Query Match          4.7%; Score 150; DB 4; Length 1942;
Best Local Similarity 20.8%; Pred. No. 0.00036;
Matches 141; Conservative 105; Mismatches 275; Indels 158; Gaps 27;

QY 22 EKIKIE-SCSERTKIFRLYSVCTVEEKALILARQNKQKAVENLNRSILTHALLMWGASY 80
DB 946 KKKLEDECSLKK-----DIDDLTLTAKVEKEKH-ATENKVKNLTEB--MAGLDE 994

QY 81 LFDKLDHFSSSTPDGVSFEGSIMGVVIHE--FSSILSK-----GGEENEVK 127
DB 995 TIAKL-----TKEKALQEAHQOTLDLQMBEDKVNTLTAKTKLEQQVDDLEGSLEQEK 1050

QY 128 LCALLE--AKHAQG--TYSDDSTLFCG-DHIKLSDESPNI FWSKLLGKNPMWKPST 182
DB 1051 LCHDLERAKRLEGDLKLAQESTMOTENDKQNEKLLKKEFEMSNLQKIEDEQALATQ 1110

QY 183 PORNRRVQ--YFEGSEASPKTDGGNAKKRKAASDDVTPRVTPDPPVDDDERKASGDHM 241
DB 1111 LQKKIKELQARIEELBEEIEAERASRAKAKQKRSLSRELEISERLEEAGGATSQAIEM 1170

QY 242 GALESPKVTIL-----QSSCKSGTGTLDGDAFLGYSNGSHISGP-----284
DB 1171 NKKREAEFQMRRLDEESTLQHEATAAALRKXGADSVAGELGQIDSLQRVQKLEKES 1230

QY 285 -----EDMLASQDWGKIPDESORRLHTVLKPKQAKL 315
DB 1231 LKMEINDLASNMETVSKANPEKMCRTLEDQLSEI---KTKEEQORLINELSAQKARL 1287

QY 316 -----CQVLHLSDACTSMGN-----FLEYIENHRIYEE---PATTFOAFQIALSWI 360
DB 1288 HTESGFSRQLDEKADAMVQLSRGKAFTQOIBELKRLQLEETKAKSTLAHALQSA-RHD 1346

QY 361 AALLVKQILSHKES-----LVANSSEAFKCSRVVEVDYIYILSCMSLFLHTQGLQ 413
DB 1347 CDLLREQYEEQEAQAKALQKMSKANSVAQWRTKYETDAI-----ORTEELE 1394

QY 414 FDCFGTNSQSVVSTKLVAESLSGATVRDEKINTK-----SMRNSSEDEECM-----T 461
DB 1395 -----EAKKLAQRLQDAEHEVAVNSKCALEKTKQRLQNEVEDLMIDVERS 1442

QY 462 EKRCSHYSTATRDIETKIIGIKKKYKQVQKLVQEHKEK-----MELLNMYADKKQL 515
DB 1443 NAACIALDKQRNFQDLVLAEWKQYBETQAELEASQKESRSLSLTELFKVKNAYEESLDHL 1502

QY 516 ETSKSVAAVIRITCSTSTQVGDLLDHNHYERKDEIKSEKNECLKSLEQM-----HE 570
DB 1503 ETLKR-ENKNLQOEISDLTEQIAEGGKHIELEKVKVQQLDHEKSELOTSLEEAASLEHE 1561

QY 571 VAKKLAEDBAACWINRIKS 589
DB 1562 EGKILRIQLE--LNQVKS 1577

```

```

RESULT 6
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11433

Query Match          4.6%; Score 147; DB 4; Length 2733;
Best Local Similarity 19.4%; Pred. No. 0.0012;
Matches 134; Conservative 117; Mismatches 250; Indels 188; Gaps 33;

QY 18 VHVVEKIKIE-----SCSERTKIFRLYSVCTVEEKALILARQNKQKAVENLNRSIL 70
DB 785 LKEIEAEKVELEKLVSSSTSELTK-----KSEEVFQLQEQINKQGLEIESL-KTVS 834

QY 71 HALLMWGASYLFDKLDHFHSSSTPDGVSFEGSIMGVVIHEPSSILSSKGGENEVKLC 130
DB 835 HEAEVHAES-LOOQLE---SSQLQIAGLEHLRELQP-KLDELQKLISK--BEDVSYLSG 887

QY 131 LLEAKHAQGTYSDDSTLFCGEHDKU-----SDESPNIFWSKLLGKNPMWKPST 182
DB 888 QLSKEEAALTKIQTBIIEQEDLIKALHTQLEMQAKEHDERI--KQLQVELCEMKQKPEI 945

QY 183 PORNRRVQY---PFGSEASPKTDGGNAKKRKAASDDV-----TDPRTVTPDPPV- 233
DB 946 GEESAKQIQIRKQALALISKEA---LKENKSLOEELSLARGTIERUTKSLADVESQV 1001

QY 234 KASGKDH---MGAL-----ESPKVIT-----LQSSCKS--SGTDGTLGDNDAFG 272
DB 1002 SAQNEKQTVLGRLLALQOERDKLITEMDRSLEENQSLSSCESLKLALGLTEDKE--- 1058

QY 273 LYSMGSHISGIPEDMLAS-----QDWGKIPDESORRLHTV-----307
DB 1059 --KLVEIESLSKSKIAESTEQEKHELOKEYEILLQSYENVNSAEARIQHVEAVRQE 1116

QY 308 -----LKPQWAKLCQVLHLSDACTSMVGNFLEVY 336
DB 1117 KOELYGKURSTANKKETEKQLQEAQEMNEEKQKRFKAK-----SQQKILELE 1167

QY 337 IENHRIYEE--PATTFOAFQIALSWIAALLVKQILSHKESLVANSSEAFKCSRVVEYD- 393
DB 1168 EENDRLRAEVHPAGD-----TAKECM---ETLSSNASMKELERVQMEYE 1210

QY 394 -----IYSLSCMSKSLFLEHTQGLQDFCGTNSKQSVVSTKLVAESLSGATVRDEKIN-- 446
DB 1211 TLSKKFQSLMSEKDSL-SEEVQDLKHQIEGNSVSKQA-----NLEATEHDNQNTV 1260

QY 447 ---TKSMRNSSEDEECM---TEKRCSHYSTATRDIETKISGIKKKYK--QVQKLVQEH 498
DB 1261 EGTQSIPGETBEQOESLMSSTRPTCSVPKSNAPAVS---KPFSSHDEINNYLQOQD 1317

QY 499 EKKMELLNMYADKKQKLETSKSVAAVIRITCSTSTQVGDLLDHNHYERKDEIKSEK 558
DB 1318 QLKERIAGLIEBKQNKKEFSQTLNEBK-NTLLSQISTKDGELKMLQEEV-TKMNLIHQOI 1375

```

QY 559 NECLSLEOMHEVAKKLADEACWINRI 587
DB 1376 QEELSRVTKLKTAEABEKKDDLEERLMNQL 1404

RESULT 7

US-09-248-796A-20275
; Sequence 20275, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 20275
; LENGTH: 630
; TYPE: PRT

; ORGANISM: Candida albicans
US-09-248-796A-20275

Query Match 4.6%; Score 146.5; DB 4; Length 630;

Best Local Similarity 20.5%; Pred. No. 0.00013;

Matches 125; Conservative 89; Mismatches 210; Indels 187; Gaps 26;

QY 108 VIHFPSILSSKGGEGNEVKLLLEAKHAQ--GTVSSDSTLFGEDHIKLSDESINFW 165

DB 40 VIKESYDKIKLSKINSIK-----ENHSEKITTHNEQKTSLKQDIKLSQDH--ELAQ 91

QY 166 SKLLGGKPNPMKYPSTPQNRKRVQYFEGSEASPKTGDGNAKRRKASDDVTDPRVTD 225

DB 92 TQLEDKENQULKELKASLEKHN-----TESATSTEEKNN--QIKELSETIKSLK-TE 139

QY 226 PVDDDERKASGDH-----MGALESPPK-----ITL 252

DB 140 LKTSGDALKSQSEYKTLTKNSDTLKLKLELEELKVKLLDQTADEKLKITERIAL 199

QY 253 QS---SCKSSGTGT-----LQNDAPGLYMSGSHISGIPEDM 287

DB 200 KSELETVKNGLSTTSELAALTITVKLSLEKEKEBELQFLSGNKSKELEDYIQKHSIDSEKL 259

QY 288 LASQDWGKIP---DESQRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYVNIENRIY 343

DB 260 KALTDELKERTKQFDSDSKKL--TELENDLTSTKKELETEKTQTSKFN-----LEERKOK 313

QY 344 EEPATTFQAFQIALSWTAALLVKQILSHKESLVANSELAFKCSR--EVDYIYVILSCM 401

DB 314 E-----IVKLNLELELLKNDNSGAKKELLEKVKLESEIEILSKELEDK 357

QY 402 KSLFLEH-----TQGLQFD-----CFGTNSKQSVWSTKL 431

DB 358 KSVMKQHDLEKQTEKKNQSLQKVTQDYSTTKLKLDELQKELDAALSFKDKPETASAKL- 416

QY 432 NESLSGATVDEKINTKSMRNSSEDECMTEKCRSHYSTAIR-----DIEKTIISGIKK 484

DB 417 TQSTSDLEAANKLINI--LISEKTEQLEKLTQKHALEKLIIDDKNADSSKTNDSLEK 474

QY 485 KYKK-----QVOKLVOHEHEKKMELNMVADKKOKLETSKVEAAVIRTCRSRTQVG 538

DB 475 ANKKVLDLETQLEKLTKELTNKL-----QKHEKL-----NNEFT 511

QY 539 DLKLLDHYERKFEDEIKSEKNECLKSLEOM---HEVAKKLA--AEDE-----ACWINRIK 588

DB 512 NLSKSHSEIKTKLDKIENKNDLVKSHKELNVEHETSTKLQVVEENKKITAKYSELK 571

QY 589 SWAAKLKVCVP 599

|||||

DB 572 KLQDKLKNCVP 582

RESULT 8

US-09-248-796A-14286
; Sequence 14286, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 14286
; LENGTH: 1648
; TYPE: PRT

; ORGANISM: Candida albicans
US-09-248-796A-14286

Query Match 4.5%; Score 145.5; DB 4; Length 1648;

Best Local Similarity 20.4%; Pred. No. 0.00074;

Matches 123; Conservative 99; Mismatches 227; Indels 155; Gaps 26;

QY 60 KAVENLNRSITHTALLMWGASYLPDKLDH-----PHSSETPDGSGVSFEQSIMDGVITHEF- 112

DB 900 KNERLUNKSLKKKIELWQTLFYIQNPNLYTLKLFNSIPYTKNTKSGQDLFOSVQLPFP 959

QY 113 ---SSILSSKGGEGNEVKLCL-LLEAKHAQGTYSDDST----- 146

DB 960 VRDSSITYHSREYFLVKLMIQLMQNDTANNNLGDITKLHLTNWIDFFTFNNHTFQRQ 1019

QY 147 ---LFGEDHIKLSD-----BESPNIFWSKLLGGKPNPMKYPSTPQNRKRVQYFEGS 196

DB 1020 HLKALLGKFVIRVDNEQVDFESDPIRYNQII--DHEMKVYGRSEKSRDISPQAAIQLP 1077

QY 197 EASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASGK-----DHM-----GA 243

DB 1078 EYSNKF--VGNLMSLRCTSDLLSML-----QKNASGNKALLQIPDHVKLICRQY 1126

QY 244 LESPKVITILQSSCKSGTGTDLGNDAFGLYSMSGSHISGIPED---MLASQDWGKIPD-E 299

DB 1127 LCQRKFPDKSDQOHLAVAGI-----FVKHYLGS-ILQVPENYGITGNNDTQKAKSKD 1180

QY 300 SQRRLHTV-----LKPRMAKLCQVLHLSDACTSMVGNFLEYVNIENRIYE- 344

DB 1181 NLRYLYRVMLQLPSMKPFNDNFKPLNE---YIMASTDTVKSIISQAIINVGBIETVYEL 1237

QY 345 ---EPATTFQAFQIALSWTAAL-LVKQILSHKESLVANSELAFKCSRVEVDYIYVILSC 400

DB 1238 HDYDDLVTQRPKLTITVNSLIQLEKSIQLQNDIITGNDQDLYK-TCVEVEKL--LISP 1294

QY 401 MKSLFLEHTQGLQFDCFGTNSKQSVVSTKLVBNSLSGATVRDEKINTKSMRNSSEDEBECM 460

DB 1295 QDMLTLDLSSVTILNPNPTQEEIYDSK-----TKTLFTA----- 1331

QY 461 TEKRCSHYSTATRD-----TEKTTISGIKKYKQVKQLVQBHEKKMELNMYADKKQ 513

DB 1332 --KRCLLYIRVQEEEDSDLLLELLISGIKPSHEQRKEIVQYEK-----AEQDI 1379

QY 514 KLETSKVEAAVIT---RITCSRTSTQVGLKLLDHYERKFEDEIKSEKNECLKSLEOMHE 570

DB 1380 SLNKSKVNTANTVNNKKSTRPYSGTSLGSLNLTYH-----ELKKXCLIEIILKLSMGE 1433

QY 571 VAKK 574

DB 1434 LTRK 1437

QY 430 LVNESLGSATVRDEKINTKSMRNSSDEDEECWTEKRSCHYSTATRDIEK-----TI 479
Db 1331 L-----KQVEDKNSFREQLBEEBAKINLEKQIATLHAQVADMKKKMEDSVGCLETA 1383
QY 480 SGIKKYYKOVOKLVQSHREK-----KME-----LLNM-----507
Db 1384 EEVKRLQKQLEGLSQSHREKVAAYDKETKRLQLOELDDLVDLDHQRQSCACNLEKKQ 1443
QY 508 -----YADKKQKLET-----SKSVEAAVIRITCSRTSTQVGDLLKLLDHNYE 548
Db 1444 KKFQDLAEBKTIISAKYABERDRAEAAREKETKALSALABEAMEQKAELERLNKQPR 1503
QY 549 RKFDTEKSEKNCEKLSLEQMHVAKVKKLADEACWINRIKSWA-----AKLV 596
Db 1504 TEMEDLMSSKDDVGKS---VHELEKSKRALEQQ--VEEMKTLQEELEDELQATEDAKRL 1558
QY 597 CVPIQS 602
Db 1559 EVNLIQA 1564

RESULT 11

US-09-538-092-1044
; Sequence 1044, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 1044
; LENGTH: 1427
; TYPE: PRP
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P30622
US-09-538-092-1044

Query Match 4.4%; Score 142.5; DB 4; Length 1427;
Best Local Similarity 20.7%; Pred. No. 0.0011;
Matches 124; Conservative 109; Mismatches 228; Indels 137; Gaps 28;
QY 31 ERTKIFRLYSVCTVEEKALILARQKQKQKAVENLRSLLTHALLMGASVLPFKLDHFS 90
Db 570 EETHQKEIKALYATEK---LSKENBSLSKLEHANKENSVDIALWKS-----KL-----616
QY 91 SETPDSCVSPFQSIMDGVIIHFFSILSSKCGEENEVK-----LCLLLEAKHAQGYSS 143
Db 617 -----ETAISHQQMBELKVSFGKGTGTAEFAELKTLQEKVRLDYQYHIEINLQNOQDS 672
QY 144 DSTLFQED-----HIKL-----SDESPNIFWSKLLGKNPMWKPYPDTPORNRK---RVQ 191
Db 673 ERAAHAKEMEARAKLVKVIKENSLARSKLDKABDQHLVEMEDTLNKLQEAIEKVK 732
QY 192 YFGSRASPKTGGGNNAKGRKKASDDVTDPRVTDPPVDDDE---RKAS---GKDHMGAL 246
Db 733 ELEVLQA-----KNEQTKVIDNFTSQLKATEEKLKLDLALRKASSEGKSEMKKLQ 784
QY 247 P-----KVITLQ-----SSCKSGTGTGLDNDAPGLYSMGSHISGIPEDMLASQDWGK 295
Db 785 QLEAAEKQIKHLEIKNAESSKASSITRELQGREL-----KLTNLOENL-----SEVSQ 833
QY 296 IPDESQRRLHTVLKPKMAKLCQVHLHSDACTSMVGNFLEYVIEHRIYEBPATTFOAFQI 355

RESULT 12

US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Hasegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; PRIOR FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1972
; TYPE: PRP
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Query Match 4.4%; Score 140.5; DB 4; Length 1972;
Best Local Similarity 18.2%; Pred. No. 0.0029;
Matches 116; Conservative 109; Mismatches 212; Indels 201; Gaps 25;
QY 22 EKIKESCERTKIFRLYSVCTVEEKALILARQKQKQKAVENLRSLLTHALLMGASYL 81
Db 967 QKLEKVTAEAKIKKL-----EDDILVMDQNNKLSK-----999
QY 82 FDKLDHFSSETPDSCVSPFQSIMDGVIIHFFSILSSKCGEENEVKLLLEAKHAQGY 141
Db 1000 -----ERKLEERISDLTTNLAE---EEEKAKNLTKLKNKH-----1032
QY 142 SSDSTLFGSDHKLSDSPNIFWSKLLGKNPMWKPYPDTPQ-----RNR 187
Db 1033 -----ESMISLEVLKKEKRSQRELEKL---KRMDGEASDLHQIADLQAIKLAQKLA 1086
QY 188 KRQVPEGSEA---SPKTGGGNNAKGRKKASDDVTDPRVTDPPVD---DDEKASGKHMGAL 244
Db 1087 KKEEELQALARLEDETSQKNALKKIRE---LEGHISDLQEDLDSERARNK---AE 1138
QY 245 ESPKVTILTSSCKSGSTGTGLDNDAPGLYSMGSHISGIPEDMLASQDWGKIPDESQRRL 304
Db 1139 KQKRDLEGELEALKTELEDTLD-----TTATQOELRAKRE-----QE 1175

```
QY 305 HTVLKPKM-----AKLCQVLHLSDACTSMVGNFLEVYIENHRIYEBPATTFOAFQIALSW 359
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1176 VTVLKALDEETRSHEAQVEMRQKHTQVVBELTEQLEQFKRAKANLDTKQTLKENAD 1235
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 360 IAALL-----VQQLSHKESLVAN-SELAFKCS-----RVEV-DYIYILSCMKSLFLE 407
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1236 LAGELRVLQGAQVEHKKKLEVLQELQSKCSGGERARAEINDKVHKLQNEVESV--- 1292
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 408 HTQGLQFCFGTNSQSVSTKLNVESLS-GATVRD-----EKINTKSMRNSSED 456
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1293 -----TGMLEAEGAKIILAKEVASLSQQLQDTQELLQETRQKLVNSTKLRLQLED 1343
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 457 BECMTEKRCSHYSTATRDIEKTSIGI-----KKYKKQVOK 492
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1344 ERNSLQEQLDDEWEAKQNLERHISTLNIQSDSKKLQDPASTVESLBEGKKRFQKEIES 1403
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 493 LVQEHHEKQELLNMYADKKQKLETSKSVEAAVIRITCSRTSTQVGDKL-IDH----- 545
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1404 LTQYVEEK-----AAAYDKLETK-----NRLQOELDDLVDLDNORQLVS 1444
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 546 ---NYERKFDEIKSEKNECLKSLEQMHEVAKKKLAEDE 580
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1445 NLEKKQKQFDQLAEBKNISKYVADRRDRAEAAREKE 1482
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

RESULT 13
US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match 4.4%; Score 140; DB 4; Length 3259;
Best Local Similarity 19.3%; Pred. No. 0.0073;
Matches 133; Conservative 117; Mismatches 251; Indels 188; Gaps 33;

QY 18 VRHVEKIE-----SCSERTKIFRLYSVCTVEEKKALILARQKRNQKAVENLRSLT 70
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1311 LXEIEAEKVELELVKVSSTSELTK-----KSEEVQLQIQINKQGLEISL-KTVS 1360
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 71 HALLMWGASYLFDKLDHFHSSETPDGVSFPEQSGIMDGVTHIEFSSILSSKGGENEVKLCL 130
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1361 HEAEVHAES-LOQKLE---SSQIAGLHRLBLOP-KLDELQKLISKK--BEDVSYLSG 1413
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 131 LLEAKHAQTYSSDSTLFGEDHDKL-----SDEESNIFWSKLLGGKNPMWKYPSDT 182
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1414 QLSEKEAALTKIQTETIEQEDLLKALHTOLEMQAKEHDERI---KQLQVELCEMKQPEEI 1471
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 183 PQRNRKRVQY---FEGSEASPKTGDGNGNKRKKASDDV-----TDPRTVTPPDV-DDER 233
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1472 GEESRAKQIQIKLQALISRKEA---LKENKSLQEELSARGTTIERLTSLADVESQV 1527
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 234 KASGDKH---MGAL-----ESPKVIT-----LQSSCKS---SGTDGTLGDNDARG 272
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
```

```
Db 1528 SAQNKEDTVLGRLLALQERDKLITEMDRSLLENQSLSSCESLKLALGLEGTEDKE--- 1584
QY 273 LYSMSHISGIPEDMLAS-----QDWGKIPDESORRLHTV----- 307
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1585 --KLVEKTESLKSSKIAESTEQEKHKELOKEYEILLQSYENVSNEAERIQHVVEAVRQE 1642
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 308 -----LKPOMAKLCQVLHLSDACTSMVGNFLEV 336
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1643 KQELYGLKLRSTEANKKETEKQLQEAEOBMEEMKERVKPAK-----SKQQKILELE 1693
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 337 TENHRIYEE--PATTFOAFQIALSWIAALLVQQLSHKESLVANSELAFKCSRVEVDY- 393
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1694 EENDRLRAEVHPAGD-----TAKECM---ETLSSNASMKEELERVKMEYE 1736
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 394 -----IYSILSCMKSLFLEHTQGLQFCFGTNSQSVSTKLNVESLSGATVRDEKIN- 446
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1737 TLSKFKQSLMEKDSL-SEEVQDLKHQIEDNVSKQA-----NLEATEKHDTQNTV 1786
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 447 ---TKSMRNSSEDEECM---TEKRCSHYSTATRDIEKTSIGIKKKYKK--QVQKLVQEH 498
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1787 BEGTQSIPEETEEOQSLSMSTRPTCSSESVPSAKSANPAVS---KDFSSHDEINNYLQQTID 1843
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 499 EKKMELLNMYADKKQKLETSKSVEAAVIRITCSRTSTQVGDKLKLDHNVVERKFDKISPK 558
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1844 QLKERIAGLBEBEKQKNEFQSTLENEK-NTLLSQISTKDGELKMLQEEV-TKQNLNLNQOI 1901
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 559 NECLKSLEQMHEVAKKKLAEDAECWINRI 587
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1902 QEELSRVTKLKETABEEKDDLEERLMNQL 1930
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

RESULT 14
US-09-949-016-7432
; Sequence 7432, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7432
; LENGTH: 1836
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7432

Query Match 4.3%; Score 138; DB 4; Length 1836;
Best Local Similarity 19.6%; Pred. No. 0.0044;
Matches 124; Conservative 99; Mismatches 217; Indels 194; Gaps 29;

QY 2 ADAFTLFGSSLNPSHDVHKVHKIEKIESCSERTKIFRLYSVCTVEEKA-----L 49
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 822 ADTVVIFDSDWNPQDLQAQRAHRIQKQKQNIYRLVTGTVVEEIERAKKMWLDHL 881
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 50 ILARONKQRKNQAVEN-----LNRSULTHALLMWGASYLFDKLDHFHSSETPDGVSF 100
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 882 VJQRMDTTGRTILNNSGRSNPNPNKEELTAILFGAEDLFKELE-----GEES 931
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 101 EOSIMDGVTHIEFSSILSSKGGENEVKLCLLEAKHAQTYSSDSTLFGEDHKLISDEES 160
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 932 EPQEND-----IDEILRLAETRENEVSTSATDELL-SQPKVANFATMEDEEL-----ER 981
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

QY 161 PNIFWSKLLGGKNPMWKYPSDTPQRNRKRVQYFEGSEASPKTGDGNGNKRKKASDDVTD 220
| | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2005, 00:58:28 ; Search time 57.6667 Seconds
(without alignments)
4160.324 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1726218 seqs, 3863311768 residues

Total number of hits satisfying chosen parameters: 1726218

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3213	99.9	2254	17 US-10-732-923-18943	Sequence 18943, A
2	591.5	18.4	1990	16 US-10-437-963-178706	Sequence 178706,
3	491.5	15.3	2630	16 US-10-437-963-172374	Sequence 172374,
4	456.5	14.2	1994	16 US-10-437-963-178698	Sequence 178698,
5	336.5	10.5	358	16 US-10-425-115-343695	Sequence 343695,
6	192.5	6.0	162	15 US-10-424-599-155808	Sequence 155808,
7	163.5	5.1	1461	17 US-10-732-923-8689	Sequence 8689, Ap
8	163.5	5.1	1465	15 US-10-369-493-5100	Sequence 5100, Ap
9	163.5	5.1	1465	17 US-10-732-923-8690	Sequence 8690, Ap
10	163	5.1	1445	17 US-10-732-923-8160	Sequence 8160, Ap
11	162.5	5.1	1739	15 US-10-341-434-61	Sequence 61, Appl

12	162.5	5.1	1739	17 US-10-732-923-8832	Sequence 8832, Ap
13	161	5.0	1015	15 US-10-369-493-3210	Sequence 3210, Ap
14	160	5.0	1738	17 US-10-732-923-8335	Sequence 8335, Ap
15	157	4.9	126	16 US-10-425-115-329771	Sequence 329771,
16	156	4.9	2552	16 US-10-437-963-129734	Sequence 129734,
17	148	4.6	1786	17 US-10-732-923-8672	Sequence 8672, Ap
18	148	4.6	1881	14 US-10-032-585-7646	Sequence 7646, Ap
19	147.5	4.6	1358	15 US-10-675-072A-43	Sequence 43, Appl
20	147.5	4.6	1813	17 US-10-732-923-8155	Sequence 8155, Ap
21	147.5	4.6	6761	17 US-10-732-923-15035	Sequence 15035, A
22	147.5	4.6	3225	16 US-10-408-765A-254	Sequence 254, App
23	145.5	4.5	1647	14 US-10-032-585-7572	Sequence 7572, Ap
24	145.5	4.5	1959	15 US-10-028-248A-36	Sequence 36, Appl
25	145.5	4.5	1959	15 US-10-028-248A-36	Sequence 36, Appl
26	145	4.5	699	16 US-10-425-115-233739	Sequence 293739,
27	145	4.5	1130	15 US-10-369-493-6751	Sequence 6751, Ap
28	144.5	4.5	728	17 US-10-732-923-8570	Sequence 8570, Ap
29	143	4.4	1379	14 US-10-205-219-5	Sequence 5, Appl1
30	143	4.4	1379	14 US-10-205-219-5	Sequence 5, Appl1
31	143	4.4	1960	15 US-10-028-248A-104	Sequence 104, App
32	143	4.4	1960	15 US-10-028-248A-104	Sequence 104, App
33	143	4.4	1960	15 US-10-028-248A-104	Sequence 104, App
34	141.5	4.4	1959	15 US-10-028-248A-107	Sequence 107, App
35	141.5	4.4	1959	15 US-10-028-248A-107	Sequence 107, App
36	141.5	4.4	3147	17 US-10-732-923-10253	Sequence 10253, A
37	141.5	4.4	3167	17 US-10-732-923-10252	Sequence 10252, A
38	140.5	4.4	1961	15 US-10-028-248A-103	Sequence 103, App
39	140.5	4.4	1961	15 US-10-028-248A-103	Sequence 103, App
40	140	4.4	906	17 US-10-732-923-6837	Sequence 6837, Ap
41	139	4.3	1806	17 US-10-732-923-8671	Sequence 8671, Ap
42	138.5	4.3	1422	17 US-10-732-923-8665	Sequence 8665, Ap
43	138	4.3	972	9 US-09-924-154-16	Sequence 16, Appl
44	138	4.3	1426	17 US-10-732-923-8310	Sequence 8310, Ap
45	137	4.3	879	17 US-10-732-923-3291	Sequence 3291, Ap

ALIGNMENTS

RESULT 1

US-10-732-923-18943
; Sequence 18943, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15 (52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18943
; LENGTH: 2254
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-18943

Query Match	99.9%	Score 3213	DB 17	Length 2254
Best Local Similarity	99.8%	Pred. No. 1.1e-236		
Matches 620	Conservative 1	Mismatches 0	Indels 0	Gaps 0
QY	1	RADAFILFGSSLNPSHDVXHKIESCSERTKIFRLYSVCTVEEKALILARQNRQNK	60	
DB	883	RADAFILFGSSLNPSHDVXHKIESCSERTKIFRLYSVCTVEEKALILARQNRQNK	942	
QY	61	AVENLNRSITHALLMWGASLYFDKLDHFHSSSTPGSGVSFEQSIMDGVITHEFSSILSSKG	120	
DB	943	AVENLNRSITHALLMWGASLYFDKLDHFHSSSTPGSGVSFEQSIMDGVITHEFSSILSSKG	1002	
QY	121	GEENEVKLCILLEAKHAQCTYSSDSTLFCEDHILKLSDEESPNIFWSKLGGKRNPMWKYS	180	
DB	1003	GEENEVKLCILLEAKHAQCTYSSDSTLFCEDHILKLSDEESPNIFWSKLGGKRNPMWKYS	1062	

QY 181 DTPQNRKRKVOYFEGSEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH 240
DB 1063 DTPQNRKRKVOYFEGSEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH 1122
QY 241 MGALSPKVITLQSSCKSGTDTLGNDAFGLYSNGSHISGIPEDMLASQDWGKIPDES 300
DB 1123 MGALSPKVITLQSSCKSGTDTLGNDAFGLYSNGSHISGIPEDMLASQDWGKIPDES 1182
QY 301 QRLHVTLPKPAKLCQVHLSDACTSMVGNFLEYVYENHRIYEBPATTFOAFOIALSWI 360
DB 1183 QRLHVTLPKPAKLCQVHLSDACTSMVGNFLEYVYENHRIYEBPATTFOAFOIALSWI 1242
QY 361 AALLVKQILSHKESLVANSELAFKCSRVEVDYIYISLSCMSLFLEHTQGLQDFCGTN 420
DB 1243 AALLVKQILSHKESLVANSELAFKCSRVEVDYIYISLSCMSLFLEHTQGLQDFCGTN 1302
QY 421 SKQSVVSTKLVSLSGATVRDEKINTKSMRNSSEDEECWTEKRCSCSHYSTATRDIEKTIS 480
DB 1303 SKQSVVSTKLVSLSGATVRDEKINTKSMRNSSEDEECWTEKRCSCSHYSTATRDIEKTIS 1362
QY 481 GIKKYYKQVQKLVQHEHEKQWELLNMYADKKQLETSKVAAVIRITCSRTSTQVGL 540
DB 1363 GIKKYYKQVQKLVQHEHEKQWELLNMYADKKQLETSKVAAVIRITCSRTSTQVGL 1422
QY 541 KLLDHNRYERKFDIEKSEKNECLSKLEQMHQHEVAKKLADEACWIRIKSWAAKLVKCVPI 600
DB 1423 KLLDHNRYERKFDIEKSEKNECLSKLEQMHQHEVAKKLADEACWIRIKSWAAKLVKCVPI 1482
QY 601 QSGNNKHFSGSSNISQAPDV 621
DB 1483 QSGNNKHFSGSSNISQAPDV 1503

RESULT 2

US-10-437-963-178706
; Sequence 178706, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 178706
; LENGTH: 1990
; TYPE: PRP
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76236C.1.pap
US-10-437-963-178706

Query Match 18.4%; Score 591.5; DB 16; Length 1990;
Best Local Similarity 26.2%; Pred. No. 1e-35;
Matches 175; Conservative 127; Mismatches 239; Indels 127; Gaps 20;

QY 3 DAFILFGSSLNPSHDVHVKHEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKQKAV 62
DB 755 DAIITCYSDNPTNDLURLQRISESQSECVPIFLRYSCTVEEKTLLAKHDHILDSNV 814

QY 63 ENLNRSLTHALLMWGASYLFDKLDHFH-----SSETPDGVSPSEQSIMDGVTHFSGSILS 117

DB 815 QNVMPVSHLSLWSGASFLNRLEREFQKHDSYKSDSDGL-----FMNVFLEPAKL 869

QY 118 SKGSENEVKLCLEAKHAQCTYSDDSTLFE--DHIKLSDSESPNI--FWSKLGGKQP 174
DB 870 TNVEASTKMNENAVISRAQSGSFYSRDIATVISEREGISAVDGLPKFTFWSNLLGGRSP 929
QY 175 WKYPSDTPQNRKRKVOYFEGSEASPKTGDGNAKRRKASDDVTDPRVTDPPVDDDERK 234
DB 930 HWQYISEPVQRRKIKQNMEDQMRIPABETDEAINKRRKIGE-----IMD----- 974
QY 235 ASGDHMGALSPKVITLQSSCKSGTDTLGNDAFGLYSNGSHISGIPEDMLASQ-- 292
DB 975 -----SSPKILPVKD-----NDAV-----LPSNSTASSHE 1000
QY 293 -----WKIPDES-----QRLHVTLPKPAKLCQVHLSDACTSMVGNFLEYVYENHRI 342
DB 1001 TSVDDTWOELGAESLQGTQKGLHTQLKPELSKLYELLELPETVKCLCEBLLDYILKNHGV 1060
QY 343 YEPPATTFOAFOIALSWIAALLVKQILSHKESLVANSELAFKCSRVEVDYIYISLSCMK 402
DB 1061 SOEPKGIILHAFNIALCWRAASLLKHKINRRSALUSVRNLNYECDEVLAEYVTEKRLIK 1120
QY 403 SLFLEHT-----QGLQDFCGTNSKQSVVSTKLVSNE--LSGATVRDEKINTKSMRNSSE 455
DB 1121 KFSRRASETSKQSQSTPWNNTSSYKQTSPLKRSDSGISCHQVTTIDGLENVSHEEAPH 1180
QY 456 D-----BECHTEKRCSCSHYSTATRDIEKTIISGIIKKYKQV--QKLVQHEHEKQWELLNMYAD 510
DB 1181 DILTBEIMILEQK-----ELISVLETHREEHVLRDELLERITEKRINLINMVS 1228
QY 511 KQKQ-----LETSKVAAVIRITCSRTSTQVGLKLLDHNRYERKFDIEKSPK 558
DB 1229 LREKNIQKQNETTLTLDHMHKQKEVAKLOETCN-----LVVLEHRLKHID--SEDR 1277
QY 559 NECLSKLEQMHQHEVAKKLADEACWIRIK-----SWAAKLVK-VPIQSGNNKHFSGSS 612
DB 1278 DATVKLIIEWFTLLLYAFNLNMRCHQNKLMQOQSTSWNKELQLKEIFLQAKSGHLDRSF 1337
QY 613 NISQAPD 620
DB 1338 DQOIPLPD 1345

RESULT 3

US-10-437-963-172374
; Sequence 172374, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 172374
; LENGTH: 2630
; TYPE: PRP
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70516C.1.pap
US-10-437-963-172374

Query Match 15.3%; Score 491.5; DB 16; Length 2630;
Best Local Similarity 25.2%; Pred. No. 7.3e-28;
Matches 176; Conservative 107; Mismatches 256; Indels 159; Gaps 20;

QY 3 DAFILFGSSLNPSHDVHVKHEKIKIESCSERTKIFRLYSVCTVEEKALILARQNKQKAV 62

Db 881 DAIISYDNNPLNDLKALQKIKIESQERFVSIFRLYPTTVEKSLVLARQGVINDNI 940
Qy 63 ENLNRSLTHALLMWGASYLFDKLDHFSSTPDGSGVFEQSDMGVIMHFFSSILSSKGE 122
Db 941 QDLRTSLKSLRLRGAAPFLSRDLVQDDHASKSEMERHFIDEVIVFLKLTIVED 1000
Qy 123 ENEVKLCLLEAKHAQTYSDSTLFGEDHDKLSDESPNIFWKLGGKNNPMKYPDSOT 182
Db 1001 STEVHRKISGKANNMGLYRNITLMEKEGIVLEDPNPAEFMLNLDGRSPHYSCISEP 1060
Qy 183 PQRNRKVOYFEGSEASPKTDGNGNAKRRKASDDV---TDPRVTPDPPVD---DERKAS 236
Db 1061 LQSRVTKSQTDW---EVNAPAEINEARKKRGKVEIMGSSKVVSDKSNDDALPDICTTS 1118
Qy 237 GKDHMGALSPKVTITLQSSCKSGTDTLGDNDAGFLYSMGSHISGIPEDMLASQDWGKI 296
Db 1119 GP---ALQ-PVDVTQKSVQSEGSILM-----1142
Qy 297 PDSQRRLHTVLPKPMKACQVHLHS-----DACTSMVGNFLEVIVIE 338
Db 1143 --STPKNLHAQMKQELSKLKVLPNTIOFGAKIAGQVLIIEVDNVTLLVEOFFEYLLN 1200
Qy 339 NRIYEPATTFQAFQIALSWIALIVKQILSHKESLVANSELAPKCSREVDVYIYSL 398
Db 1201 NHVVVQEPKYIFALNIALCWRVASIHNFVKDHKESLALAEKRLKYECNEBELARLVYDSL 1260
Qy 399 SCMKSLFLEHTQGLQFCFGTNSKQSVVSTKLNVNESLSGATVRDEKINTK-----448
Db 1261 ---KRPEPKAGATGSCNST---SVETKPSQOETSN-ILRNDHIFPKQMDLHDNFM 1312
Qy 449 -----SMRNSSEDECM---TEKRCSHYSTATRD---TEKTSIGIKKKYKQVQK 492
Db 1313 NGALQEGFVAAQWVSEBELIAPVPGTHMEC-HFSTDELDPDIVEKRLINLIDNVFSVREYR 1371
Qy 493 LVQEHKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGLKLLDHNRYKFD 552
Db 1372 IFDQKQSQISELEKYQNTKARUKT-----VCN-----LVLEHICRSHAD 1411
Qy 553 EIKSEKNECLSLQFQMEHVAKKLADEACWINRIKS-----WA-----AKLKVCVPTQSG 603
Db 1412 --VETRNDTIKQVQWFTMLMYAFLEHMLRHLQHSKLSQNTWAEERQLKELCLEAKSG 1469
Qy 604 -----NNKHFGSS 612
Db 1470 QLDHTFDQIALPDSNFMQEFTHLKEQSSNSHVSGSA 1507

RESULT 4

US-10-437-963-178698
; Sequence 178698, Application US/10437963
; Publication No. US20040123343A1

GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 178698

; LENGTH: 1954

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1.pap

US-10-437-963-178698

Query Match 14.2%; Score 456.5; DB 16; Length 1954;

Best Local Similarity 24.0%; Pred. No. 2.3e-25;

Matches 160; Conservative 118; Mismatches 229; Indels 159; Gaps 24;

Qy 3 DAFILGSSINPBDVKKHVEKIKIESCSERTKIFRLYVCTVEEKALILARQNRQKAV 62
Db 755 DAIISYDNNPLNDLRVLQRISESQSECVPFRLYSSCTVEKTLIL-----803
Qy 63 ENLNRSLTHALLMWGASYL---FDKLDHFSSTPDGSGVFEQSDMGVIMHFFSSILSSK 119
Db 804 ---RM---IIFATATSRILEEFQKD-YSSKSEDDGL-----FMNVFLEPAKLTSTN 851
Qy 120 GGENEVKLCLEAKHAQTYSDSTLFGEDHDKLSDESPNIFWKLGGKNNPMW 176
Db 852 VEASTKMNNAVIGRAQSGSFYSRDIAVISEREIGISAVDGLPKFTFWNLLGGRSPHW 911
Qy 177 KYPSDTPQRNRKVOYFEGSEASPKTDGNGNAKRRKASDDVTPDPPVDDDERKAS 236
Db 912 QYISEPVQRNRKIQNMEDQMRIPABETDEAIMKRRKIGE-----IMD-----954
Qy 237 GKDHMGALSPKVTITLQSSCKSGTDTLGDNDAGFLYSMGSHISGIPEDMLASQD---292
Db 955 ---SSPKILPVKD-----NDAV-----LPENSTASSSHETS 982
Qy 293 ---WGKIPDES---QRLHTVLPKPMKACQVHLHSDACTSMVGNFLEVIVENHRIYE 344
Db 983 VDDTWQELGAESLQGTQKGLHTQKPELSKLYELLELPETVKCLCEBLLDYILKNHQ---1039
Qy 345 EPATTFQAFQIALSWIALIVKQILSHKESLVANSELAPKCSREVDVYIYSLCKWSL 404
Db 1040 -----CWRAASLLKHKINRRRESLALSVRLNAYECDEVLAEVYVEKRLILKCK 1086
Qy 405 FLEHT---QGLQFDCFGTNSKQSVVSTKLNVNES---LSCATVRDEKINTKSMRNSSE-456
Db 1087 FSRASSETSKQSGSTPWNNTSSYKQOTSPKLRSDGSGICHQVTTIDGLENVSHHEAHPDI 1146
Qy 457 ---EECHTEKRCSHYSTATRDIEKTSIGIKKKYKQV---QKLQVHEHEKKMELLNMYADKK 512
Db 1147 LTEEMILEQK-----ELISVLETHREHVLRLDELLERITKRLINLNMVFSLR 1194
Qy 513 QK-----LETSKSVAAVIRITCSRTSTQVGLKLLDHNRYKFDIEKSEKNE 560
Db 1195 EKNIDQKQNETTLLDMHKQKQVAKLQETCN-----LVVEHLRKGHID--SEDRDA 1243
Qy 561 CLAKSLQMEHVAKKLADEACWINRIK-----SWAAKLKVC-VPIQSGNNKHFGSSNI 614
Db 1244 TVKLIIEWFTLLLYAFNLNMRQCHNKLKMQOOSTSMNKLQKELQKIFLQQAAGSHLDRSPDQ 1303
Qy 615 SQNAPD 620
Db 1304 QIPLPD 1309

RESULT 5

US-10-425-115-343695

; Sequence 343695, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 343695

; LENGTH: 358

; TYPE: PRT

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(358)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_76611C.1.pep
US-10-425-115-343695

Query Match      10.5%; Score 336.5; DB 16; Length 358;
Best Local Similarity 31.0%; Pred. No. 3.3e-17;
Matches 102; Conservative 58; Mismatches 120; Indels 49; Gaps 12;

QY 3 DAFILGSSLNPSHDVHKVHEKIKIESCSERTKIFRLYSVCTVEEKALILARONKQKAV 62
Db 47 DAIILYCSDNWPNMDLRLQRVSMESQSEVPVIFRLYSFTVEEKALILAKHDHILDSNI 106
QY 63 ENLNRLSLTHALLMWGASYLFDKLDHF--HSSETPDGVSFQSGIMDGVTHEFSSLSK 120
Db 107 VNITPSLSCHLSWGASFLNRLLEELQOHSY----SNVSGDELFMNDVDFLTKLLSKV 162
QY 121 GEENEVKLCLEAKHAQGTYSSTLFGEDHILKLSDESPNI--FWSKLLGGKPMWK 177
Db 163 ELRTSGNTAIQAYLCGSFYRAIVVAGEREGIPSDGLPKFWAYWVSLNGRSPQWQ 222
QY 178 YPSDTPQNRKRVQVPEGSEASPKTGDGNAKRRKKASDDVTDPRVTPDPPVDDDERKASG 237
Db 223 YISEPVQSRRKI---NNMEQQLKNTD---KQLKITTEDEARV-----KERRIG 267
QY 238 K-DHMGALSPKVTITLQSSCKSGTGTLDGNDAGFLYSGMGSIHGIPEDMLASQDWGKI 296
Db 268 EIMDSASIDSP-----GKNKDTILPGNNT---PPSSH-----QISVEDTWQBL 307
QY 297 P----DESORRLHTVLKPKMAKLCQVLHL 321
Db 308 ERSNLHATQGLHVQLKPEISKDYKLLQL 336

RESULT 6
US-10-424-599-155808
; Sequence 155808, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 155808
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1.pep
US-10-424-599-155808

Query Match      6.0%; Score 192.5; DB 15; Length 162;
Best Local Similarity 29.2%; Pred. No. 1.2e-06;
Matches 57; Conservative 32; Mismatches 71; Indels 35; Gaps 4;

QY 337 IENHRIYEEPAITFOAQIALSWIAALLKQILSHKESLVRANSELAFKCSRVVDYIYS 396
Db 1 MNLYNREPFPSISQAFSLCWTAAASLLKHLDPIASLIQ----DLNFECKEEDYDICS 57
QY 397 ILSCKMSLFEHTQGLQFCFGTNSQSVVSTKLVNESLSGATVRDEKINTKSMRNSSED 456
Db 58 MLCCLLKFIPLYRTGNVHTDGF-----TKASGFSNRA 88
```

```
QY 457 EECMTEKRCSHYSTATRDIEKTIISGIIKKYKQVQKLVQEHHEEKQKQELLNMYADKKQKLE 516
Db 89 YSCTGVAR--EVELFKDMSKSIKEIQKCKEKKLKLHLOEBEQRRLAATEEBEKANPD 146
QY 517 TSKSVEAAVIRITCS 531
Db 147 ERYNIESADIR-SCS 160

RESULT 7
US-10-732-923-8689
; Sequence 8689, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgeton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8689
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-732-923-8689

Query Match      5.1%; Score 163.5; DB 17; Length 1461;
Best Local Similarity 17.6%; Pred. No. 0.0045;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAPILGSSLNPSHDVHKVHEKIKIESCSERTKIFRLYSVCTVEEKALILARONKRONKA 61
Db 799 ADTVIIFDSDWNPQNDLQAMSPAHRIQGTKTWIVRLVTKGSVEE-----EI 845
QY 62 VENLNRLSLTHALLMWGASYLFDKLDH-----FHSSETPDGVSFQSGIMDGV 108
Db 846 VERAKRKLIV-----LDHLVIQRMDDTKTVLSKNATASGVPDPKQELSAI 891
QY 109 IHERPSSI--LSSKGGEENEV-----KLCILLLEAKHAQGTYSSTLFGEDHILKLSDES 160
Db 892 L-KGAVELFKKEGEBQEPEDVIDIRILMGATREABEEVWKENELLSFKYANFAIDEE 950
QY 161 PNIF-----WSKLLGGKPMWKYPSTDPQRNRKV-----Q 191
Db 951 KDIAAATDEWAAII-----FEEDNRILEEERMKELAEMLAPRQRKQIPQ 997
QY 192 YPEGSEA-SPKTGDGNAKRRKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRFIKSFRKFSMPLARLEEIAQDAELE 1057
QY 235 ASGKDHMGALSPKVTITLQSSCKSGTGTLDGNDAGFLYSGMGSIHGIPEDMLASQDWG 294
Db 1058 EHSDEMKKL-----VESLSEACKKAAD--FDSNENKNG-----DAG 1092
QY 295 KIPDESQ-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEXY 336
Db 1093 AAESEKKDIERKFKPHTCDVNLKQIERSHAELKP---LHEILKSEETKTS----- 1139
QY 337 IENHRIYEEPAIT-----FQAQIAL-----SWIAALLVKQI-LSHKESL 375
Db 1140 -----FKPPANAKLQKQWDVMSRPDSDSALLGVWKYGYGSWEAKKMDPTLGLADKIFI 1193
QY 376 VRANSELAFKCSRVVDYIYSILSCWKSILFEHTQGLQFCFGTNSQSVVSTKLVNESL 435
Db 1194 KDKTKKPGQNLQVADVLLKLM-----SKOKVKTTE----- 1225
QY 436 SGATVRDEKINTKSMRNSSEDEECMTEKRCSHYSTATRDIEKTIISGIIKKYKQVQKLVQ 495
Db 1226 -----KKERKRKADDVPVGPEKKKGRHTNNVPOEGE-----KKKEKK----- 1262
QY 496 EHEEKKMELLNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFEIK 555
```

Db 1263 --BEKNSSSLK---DQALLSIDKSLYGGALESSAKPFLECVKLCMPVHKYMKLKEAQ 1317
QY 556 SEKNECLKS-----LEQMEHVAKKLADEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKNQADEAKYFLRGDSFLENLETLIKKPKTNIRKWNLYLWIFLCKFTLREP 1371

RESULT 8

US-10-369-493-5100
; Sequence 5100, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5100
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5100

Query Match 5.1%; Score 163.5; DB 15; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0046;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAPILFGSLNPSHDVHKVEKIKIESCSRTKIFRLYSVCTVEEKALILARQNKA 61
Db 799 ADTVIIFDSQNDLQAMRAHRIGQTKVNIYRLVTKGSVEE-----EI 845
QY 62 VENLNESLTHALLMWCASYLFDKLDH-----FHSSETPDGVSFQSDIMDGV 108
Db 846 VERAKRLV-----LDHLVIQRMDDTKTVLSKNATASGVPPDKQELSAI 891
QY 109 IHEFSSI--LSSKGGGENEV-----KLCLLEAKHAQTYSSDSTLFG--DHIKLSDSES 160
Db 892 L-KFGAVELFKKEGEEQPEVDIDRLMGAETREAEVEVMKENELLSFKYANFAIDEE 950
QY 161 PNIF-----WSKLLGGKNPMKYPSTDPQNRKRV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDNRILIEERMKELAEWNLAPRQKQPIQ 997
QY 192 YFGESEA-SPTGDDGNNAKRRKASDDVTPRV-----TDPVDDDERK 234
Db 998 WVDDDDGDDDEEDDTGKKKKKAVGNFTPEIKRFKFRKFSMPLNRLLEEIAQDAELE 1057
QY 235 ASCKDHMGALSPKVTILOSCKSSGTDGTLGDNAFLYSGMSHISGIPEDMLASQDMG 294
Db 1058 EHSTDENKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
QY 295 KIPDESQ-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYV 336
Db 1093 AAESEKKDIERKFKFHTCDVNLKQIERSHAELXP-----LHEILKSEBTKS-----1139
QY 337 IENHRIYEEPAAT-----FOAFQIAL-----SWIAALLVKQI--LSHESL 375
Db 1140 -----FKPPANAKLQKGDWDSRPDDSSALLGVWKYGYGSWEAKMDPTLGLADKIFI 1193
QY 376 VRANSELAFCSSVEVDYIYISLSCMKSLFLEHTQGLQFCGFTNSKQSVVSTKLVNESL 435
Db 1194 KDKTKPQGNLQVRVDYLLKLM-----SKDKVKTTE-----1225
QY 436 SGATVRDEKINTKSMRNSSEDEECWTEKCRSHYSTATRDIEKTIISGIIKKYKQVQKLVQ 495

Db 1226 -----KKERKKKADDDVEVGPEKKRKRHTNNVPQGEK-----KKBKX-----1262
QY 496 EHEKKMELINMADKKQKLETSKSVAAVIRITCSRTSTQVGDLLKLDHNYERKFEDEIK 555
Db 1263 --BEKNSSSLK---DQALLSIDKSLYGGALESSAKPFLECVKLCMPVHKYMKLKEAQ 1317
QY 556 SEKNECLKS-----LEQMEHVAKKLADEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKNQADEAKYFLRGDSFLENLETLIKKPKTNIRKWNLYLWIFLCKFTLREP 1371

RESULT 9

US-10-732-923-8690
; Sequence 8690, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8690
; LENGTH: 1465
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-732-923-8690

Query Match 5.1%; Score 163.5; DB 17; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.0046;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAPILFGSLNPSHDVHKVEKIKIESCSRTKIFRLYSVCTVEEKALILARQNKA 61
Db 799 ADTVIIFDSQNDLQAMRAHRIGQTKVNIYRLVTKGSVEE-----EI 845
QY 62 VENLNESLTHALLMWCASYLFDKLDH-----FHSSETPDGVSFQSDIMDGV 108
Db 846 VERAKRLV-----LDHLVIQRMDDTKTVLSKNATASGVPPDKQELSAI 891
QY 109 IHEFSSI--LSSKGGGENEV-----KLCLLEAKHAQTYSSDSTLFG--DHIKLSDSES 160
Db 892 L-KFGAVELFKKEGEEQPEVDIDRLMGAETREAEVEVMKENELLSFKYANFAIDEE 950
QY 161 PNIF-----WSKLLGGKNPMKYPSTDPQNRKRV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDNRILIEERMKELAEWNLAPRQKQPIQ 997
QY 192 YFGESEA-SPTGDDGNNAKRRKASDDVTPRV-----TDPVDDDERK 234
Db 998 WVDDDDGDDDEEDDTGKKKKKAVGNFTPEIKRFKFRKFSMPLNRLLEEIAQDAELE 1057
QY 235 ASCKDHMGALSPKVTILOSCKSSGTDGTLGDNAFLYSGMSHISGIPEDMLASQDMG 294
Db 1058 EHSTDENKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
QY 295 KIPDESQ-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYV 336
Db 1093 AAESEKKDIERKFKFHTCDVNLKQIERSHAELXP-----LHEILKSEBTKS-----1139
QY 337 IENHRIYEEPAAT-----FOAFQIAL-----SWIAALLVKQI--LSHESL 375
Db 1140 -----FKPPANAKLQKGDWDSRPDDSSALLGVWKYGYGSWEAKMDPTLGLADKIFI 1193
QY 376 VRANSELAFCSSVEVDYIYISLSCMKSLFLEHTQGLQFCGFTNSKQSVVSTKLVNESL 435
Db 1194 KDKTKPQGNLQVRVDYLLKLM-----SKDKVKTTE-----1225
QY 436 SGATVRDEKINTKSMRNSSEDEECWTEKCRSHYSTATRDIEKTIISGIIKKYKQVQKLVQ 495

```
Db 1226 -----KKERKRKADDPVGPPEKKKRHTNNVPQGEK-----KKKEKK----- 1262
QY 496 EHEEKMELLNMYADKKLETSSKVEAAVIRITCSRTSTQVGDLLKLLDHNHYERKFEIK 555
Db 1263 --EKNSSSLK--DQALLSIDKSLYGLGALEDSAKPFLCVKLCMPVHKYMKKLKEAQ 1317
QY 556 SEKNELCKS-----LEQHEVAKKLADEACWNRKLSWAAKUKVCVP 599
Db 1318 EAKNOADEAKYLTRLGDSFLENLETLIKKPKTNIRKKNYLNWFLCKFTLEP 1371

RESULT 10
US-10-732-923-8160
; Sequence 8160, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8160
; LENGTH: 1445
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-732-923-8160

Query Match 5.1%; Score 163; DB 17; Length 1445;
Best Local Similarity 18.4%; Pred. No. 0.0049;
Matches 146; Conservative 111; Mismatches 211; Indels 326; Gaps 36;

QY 2 ADATILFGSLNPSHDVKEK-IKIESCSERTKIFRLYSVCTVEKALILARQNKRONK 60
Db 551 ANVILFDFTWNPANDLQADIRAYRIGQCRD-VKFRLLISLGTVE-IMYLKQVYKQULH 608
QY 61 AV-----ENLNR-----SLTHALLMWGASYLF-----DKLDH----- 87
Db 609 CVVVGSENAKRYFEAVQGSKEHGFVGHNFKLRSQGSCLTRDILEREQVGEAGIMTA 668
QY 88 -----FHSSETPD-----SGVSFEQIMQGVTHE----- 111
Db 669 TTLLKGEASAKLETPRDPCQEPPTNVCELYSDMSDEESVGRVAKVHKHVGSPSRTPGSSAQ 728
QY 112 -----FSSILSSK-----GGENEVKLCILLEAKHA--QGTVSSDSTLF 148
Db 729 LTLLQGFSLKFEAKCESVQDGNPNVSSDGSDEQPMCLSAKARQAACQKTWDSVCT-- 786
QY 149 GBDHIKLSDESPNIFWSKLLGKNPMWKYPSPDTPQNRKRQVYFEGSEASPKT----- 202
Db 787 -SEH-----QKSDNI-----QTPDEKCVSD--KSEKLEQNVYSSSDDETKCHSTA 829
QY 203 -----GDGNNAKRRKASDDVTP--RV-----TDPVP--DDDERK 234
Db 830 GHHCMQGGTGESDSDVIFPTOPTQRPKNPTICCKLLGSESDSESTEDPVKNHDDGRQ 889
QY 235 ASGK-----DHMGALSPKVITLQSSCKSGTGTLDGNDADFGLYMGSHISGIPEDMLA 289
Db 890 NSGRNGPNVSNFLYLEN---MTLKSVRKRKGTDDISDESDDIMF-----PKSRR 937
QY 290 SQDWG---KIPDESQRRLHTVLKPKMAK-LCQVHLSDACTSMVGNFLEYVNIENHRIYEE 345
Db 938 KQRATASLKLKSKKENKRKHNSPITAKDNTQVCEADGDCSSQ-----VIEDFSSDD 990
QY 346 PATTFQFOIALSWIAALLVKQI-----LSHKSLSVRANSELAPKCRSEVVDIYISLSCW 401
Db 991 SSS-----VSHVSFTKLNRHAEFTVKDGVQ-----EVAYIH----- 1020
QY 402 KSLFLEHTQGLQDFCGTNSKQSVSTKLWNESLSGATVRD----- 442
Db 1021 -----SNQNVIGSSRAENHMSRWATRDVPELKQFQLPANVAVCSS 1061
```

RESULT 11

```
US-10-341-434-61
; Sequence 61, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-61
```

```
Query Match 5.1%; Score 162.5; DB 15; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0069;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

QY 2 ADATILFGSLNPSHDVKEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVVIFDSMNPNQDLQARAHRIQKQKVNIYRLVTGTVVEEIIERAKKQMLDHL 938
QY 50 ILARQNKQKNAVEN-----LNRSILTHALLMWGASYLFDKLDHFSHSETPDGVSF 100
Db 939 VIQRMDTTGRTILENNSGRSNPNFKBELTAILKFGAEDLPKELE-----GBES 988
QY 101 EQSIGMDVTHEPSSILSSKGGEEVVKLLLEAKHAQGTYSDDSTLTFGEHILKLSDEBS 160
Db 989 EPQEMD-----IDEILRLAETRENEVSATDELL-SQFKVANFATMEDEEL---EER 1038
QY 161 PNIFWSKLLGKNPMWKYPSPDTPQNRKRQVYFE-----GSEASPKTG 203
Db 1039 PHKDWEI1-----PEQRKKVBEERQKELEIYMLPRINSSTKKAQTND 1085
QY 204 DGNNAKRRKASDDVTPDPRVTPDPPVDDDERKASGDKHMGALSPKVITLQSSCKSGTGD 263
Db 1086 DSDTESKQQAQSSASETE-----SDDDKK-----PKRRGRPRVRKDLVGE 1130
QY 264 TLIDGN---DAFGLYSMGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLKPKMAKLC 316
Db 1131 FTDAIRIRFIKAYKF-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174
QY 317 QVHLSDACTSMVGNFLEYVNIENHRIYEEPATTFQFOIALSWIAALLVKQLSHKESLV 376
Db 1175 ELIH--NSCVSAMQYEEQLKENASEGKPGKR-RGPTIKISGV-QVNVKSIQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEVDIYISLSCWKSLEHTQGLQDFCGTNSK 422
```


Db 1231 MLHKSIPVDPEEKKKYCLTCRVAAHFDVWGVDDSRLLGLGYHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMRNSSE-----DDECMTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLRKLGLEKKGAVTGGEEAKLKR 1342
QY 465 -----CSHYSTATRDIETKISGIIKKYKQVQKLVQEHBEKK 501
Db 1343 KPRVKKENKVPRLKEEHGIELSPRHSNDPSEGEVKDDGLEKSPMKKKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KOKLETSKVEA-----AVRITCSTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKDKKPKSGDAKSSSKRSQGPVHITAGSEPVPIGEDED 1461
QY 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLDDQETFTICKERMMPVKALKQDKPKGLNVQGLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMHEVAKKLADEACWINRKSAAKLVQCV 598
Db 1522 DQSHIKLWRNLWIFVSKFTFDPARKLHLYKMAHKRSQEE-----BQKKDDV 1572
QY 599 PIQSGNKKHP-----SGSSN---ISON 617
Db 1573 ---TGGKKPFRPEASGSSRDSLISQS 1595
RESULT 12
US-10-732-923-8832
; Sequence 8832, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8832
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-732-923-8832
Query Match 5.1%; Score 162.5; DB 17; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.0069;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
QY 2 ADAPILFGSSINPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVIFDSQWNPQNDLQARAHRIQKQKQVNIYRLVTKGTVEEHLIERAKKXVLDHL 938
QY 50 ILARONKRONKAVEN-----LNRLTHALLMAGASYLPDKLPHFSSETSPDSGVSF 100
Db 939 VIQRMDTTGRTILENNSGRNSPNFNKEELTALFKGAEDLPKELE-----GEES 988
QY 101 EQSIMDGVHIEFSSILSSKGGENEVKLCLLLEAKUAQGTYSDDSTLFGEDHKLSDERS 150
Db 989 EPQEMD-----IDEILRLAETRENEVSTSATDELL-SQKVFANFATMEDEEL-----EER 1038
QY 161 PNIFWSKLLGGKPMKVPSTQPNRKRVOYFE-----GSEASPKTG 203
Db 1039 PHKWDREII-----PEQRKQVEEEROKLEEELIYMLPRISSTTKAQTND 1085
QY 204 DGGNAKRRKASDDVTDPRVTPDPPVDDDERKASGDHMGALSPKVITLQSSCKSGTGD 263
Db 1086 DSDTESKRAQSSASESETED---SDDDKK-----PKRGRPRSVRKDLVEG 1130
QY 264 TLDGN-----DAFLGLSMGSHISGIPEDML--ASQDWKIPDPSQBRLLTVLKPMAKLC 316

Db 1131 FTDAETRRRIKAYKFP-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174
QY 317 QVLHLSDACTSMVGNFLEYVNIENHRIYEEPATTFQAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIH--NSCVSAQMEVEEQLENASEGKGPGKR-RGPTIKISGV-QVNVKSIHQHEEPE 1230
QY 377 RANSEL-----AFKCSRVEVDYIYSILSCMKSLFLEHTQGLQPCFGTNSK 422
Db 1231 MLHKSIPVDPEEKKKYCLTCRVAAHFDVWGVDDSRLLGLGYHGYG-NWELIKTDPE 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMRNSSE-----DDECMTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPGKQLQTRADYLLKLRKLGLEKKGAVTGGEEAKLKR 1342
QY 465 -----CSHYSTATRDIETKISGIIKKYKQVQKLVQEHBEKK 501
Db 1343 KPRVKKENKVPRLKEEHGIELSPRHSNDPSEGEVKDDGLEKSPMKKKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KOKLETSKVEA-----AVRITCSTSTQVGD--- 539
Db 1402 EKQMSRKDKGDKERKSKDKKPKSGDAKSSSKRSQGPVHITAGSEPVPIGEDED 1461
QY 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLDDQETFTICKERMMPVKALKQDKPKGLNVQGLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMHEVAKKLADEACWINRKSAAKLVQCV 598
Db 1522 DQSHIKLWRNLWIFVSKFTFDPARKLHLYKMAHKRSQEE-----BQKKDDV 1572
QY 599 PIQSGNKKHP-----SGSSN---ISON 617
Db 1573 ---TGGKKPFRPEASGSSRDSLISQS 1595
RESULT 13
US-10-369-493-3210
; Sequence 3210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3210
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3210
Query Match 5.0%; Score 161; DB 15; Length 1015;
Best Local Similarity 20.1%; Pred. No. 0.0042;
Matches 129; Conservative 109; Mismatches 253; Indels 150; Gaps 28;
QY 17 DVKHVEKIKIESCSERTKIFRLYSVCTVEEKLILARONKRONKAVENLNRLTHALLMW 76
Db 225 DIEHWRQVQTDLKKKEDSLRL-----QEEA---TRKTEQADOMESLNTLN----- 269
QY 77 GASFLPKDLHDFHSSSTPDSGVSEFQSIMDGVHIEFSSILSSKGGENEVKLCLEAK- 135
Db 270 -----DINDTAALAGVPIKGGADPHDGTATILEVDFNINLY--EDTEORLVQTEAKD 321
QY 136 HAQGTYSSTSTLFGEDHKLSDERS-----PNIFWSKLLGGKPMKVPSTQ 182

Db 322 KAAAAKESATRSRSETIOKLTSNNKLEQRDEQKAKADAFYQKL---KKKGWLEPTED 378
Qy 183 P-----QRNRKRVQYFEGSEASPKTGDGNNAKR-KKASDDVTDPRV---TDPVDD 230
Db 379 PSASLALDQSLKQDQWSESEKNAVNQKITNLKSIIEANVALKDKMLQVVERDSLVDL 438
Qy 231 DERKASGDHMGALLES--PKVITLQSSCKSGTGTDLGNDAFLGYSMSHISGIPEDML 288
Db 439 QORQE-----GDIKSLNQQLDLDKOKLASKA-----SAGADLKDIQSLRL 478
Qy 289 ASQDWGKIPDESQRRLHTVLKPKMAKLCQVHLSDACTSMVGNFLVYVHENHRIEPEAT 348
Db 479 ENK---SLEDQQR-----VQAEKVLQQLSOTKARLEKVTTLKNTASQKMDLETQRN 530
Qy 349 TFOAQIALSWIAALAVKQI-LSHKESL-----VRANSELAFKCSRVV----- 391
Db 531 EWSKAKVLEEEISRLKEVDLARKNSAAPTIVEIKPEPNIAESITRVQLMKEQKPOLP 590
Qy 392 ----DIYIILSCMKSL-----FLEHTQGLQPCFGTNSQSVSTKLVNESLSGATVRDE 443
Db 591 LTARDPVQGLVAKLKDAQGHKDLASEDVIRECIG-----LLNRVLK-BPVFDE 638
Qy 444 KIN-----TKSMRNSSEDEE--CMTEKRCSHYSTATRDIEKTIISGIIKKYKQVO 491
Db 639 KOKEMLOTQNELNEALRGKEDRQSAQLOSIIDKYEKESLEKEV-----KVHKEIK 694
Qy 492 KLVQHEEKKMELLN-----MYADKQKLETSKVSEAAVIRITCSRTSTQV---G 538
Db 695 KLKQETITEKTSLANQKQERDMLKESYEQIKNLNADHKKAAAEKLVKHQNELTQLRKDG 754
Qy 539 DLKLDHNYERKPEDEKSENECKLSLEQWHEVAKKKLAED 579
Db 755 DLK--ETNLQKLDTLRQONESERNRLQADYAEKAKLTKD 793

RESULT 14

US-10-732-923-8335
; Sequence 8335, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8335
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-8335

Query Match 5.0%; Score 160; DB 17; Length 1738;
Best Local Similarity 17.4%; Pred. No. 0.011;
Matches 133; Conservative 102; Mismatches 242; Indels 286; Gaps 28;

Qy 2 ADAPILPGSLNPSHDVGHVKEKIKIESCSERTKIFRLYSVCTVEEKALILA----- 52
Db 1041 ADTVIFDSQWPNQDQMSRAHRIGQQEVNVIYFVTSKSEVEILERAQKRWLDHL 1100
Qy 53 -----RQNKRONKAVENLNRLTHALLMWGASYLFDKLDHPHSSETPDGVSFEQS 103
Db 1101 VIQKLAEGLEKREKTKGSPDKNELSAILRFGAELPKE----- 1141
Qy 104 IMDGVTHFSSILSSKGGNEVKLCIL-----LEAKHAQGTSSDSTLFGEDH 152
Db 1142 -----DKNDESKRLLSMOIDEILERAQVEEKH-----TDETHELLGAPK 1184
Qy 153 I-KLSDSESPNIFWSKLLGGKPMWKPSPDTPQNRKRQVYFEGSEASPKTGDGNAKRR 211
Db 1185 VANFCNAEDGGSFWSRWIKPDSVVTAEALAPRAARNTKSYDPSPHD-----RTSKRK 1238

Qy 212 KKASDDVTDPRVDDDERKASGDHMGALLESFKVITLQSSCKSGTGTDLGNDAF 271
Db 1239 KKGSE-----PPBHTERSQKRRTKTEYFVPSTPLLEGTSAQVRGWSYGNLPRKDAQ 1288
Qy 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLCQV 318
Db 1289 RYRTVMKFGNHQWACTAEBVGVE-----AAPEEAQVELFDA----- 1328
Qy 319 LHLSDACTSMV--GNF-----LEYYVIEHNHRIEPEATTFQ 351
Db 1329 --LIDGCKESVETGNFEPKGPVLDFFGVVPKANELLKRVQGLQLLSKRISRYNDPTSQFR 1386
Qy 352 AFQIALSWIA-----ALLVKQLSH---KESLVRANSELAF--KCSRVEV 391
Db 1387 ----VLSYKPSNWSKGCWGNQIDARLLGLYHGFNGWEKIRLDESGLTKKIAPVEL 1442
Qy 392 DVIYIILSCMKSLFLEHTQGLQPCF-----GTNSKQSVSTKLVNESLSG--ATVRDE 443
Db 1443 QHHTFLPRAPMLKERATALLEMELAAAGGKVTNAKARKNKKVVDNLINQFKAPARDR 1502
Qy 444 K-----INTK-----SMRNSSE-----DDE 458
Db 1503 RCKSGPANVSLSTKDGPRKTKQARPLVKEGEMSDGGEVYEQPKQKMWECEDVLADE 1562
Qy 459 CMTEKRCSHYSTATRDI--EKTISGIIKKYK---KQVQKLVQHEEKKMELNMYADKKQ 513
Db 1563 IKTLGRLOQLTTSADLPKRVLFKIRRYLEILGRIDAIVLEHEE-----DLYKQDRM 1616
Qy 514 KLETSKSVAAVIRITCSRTSTQVGDLLDHNRYERKPEDEKSENECKLSLEQWHEVAK 573
Db 1617 TWRL-----WNYVSTSTNLSGDR-----LNQIYSKUL 1643
Qy 574 KKLAEDEACWINRIKSWAAKLVCPFIQSGNNKHFGSSNSIQ 616
Db 1644 QKEEEEGV-----GPSHLNGSRNFQR 1665

RESULT 15

US-10-425-115-329771
; Sequence 329771, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 329771
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1.pap
US-10-425-115-329771

Query Match 4.9%; Score 157; DB 16; Length 126;
Best Local Similarity 35.4%; Pred. No. 0.00044;
Matches 40; Conservative 18; Mismatches 47; Indels 8; Gaps 3;

Qy 283 IPEDMLASQDWGKIPD---ESQ-----RLHTVLKPKMAKLCQVHLSDACTSMVGNFLY 335
Db 10 LEMSTPSADLHLLPETGVENMRTPKSSPHALKRELSKLVKLPDNLVFLANGSLEY 69
Qy 336 VIENHRIEPEATTFQFOIALSWIAALVQI-LSHKESLVRANSELAFKCS 387
Db 70 LLNHLVVRPRISILHAFNIALCWRASAPLKYTELDHRESLALASDGLNYECN 122

Search completed: July 12, 2005, 01:07:29
Job time : 60.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 11, 2005, 09:40:15 ; Search time 7459.33 Seconds
(without alignments)
4033.965 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILFSSLNPSHDVXH.....SGNNKHFGSSNISQAPDV 621

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2 1/USPTO spooll/ROBINSON929/runat 11072005 102152 29171/app query.fasta_1.2325
-DB=GenEmbl -QPWT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=ROBINSON929 @CGN 1 1 12596 @runat 11072005 102152 29171 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3210	99.8	6554	8 AF213627	Arabidopsis
2	3210	99.8	6571	6 AX063583	Sequence
3	2977	92.6	10329	8 AF213628	Arabidopsis
4	2977	92.6	120965	8 AC026875	Genomic s

5	2974	92.5	10329	6	AX063582	Sequence
6	1033	32.1	895	8	ATM552638	Arabidopsis
7	522.5	16.2	4027	8	AK068790	Oryza sat
8	494.5	15.4	467	6	AX063609	Sequence
9	472	14.7	515	6	AX063611	Sequence
10	442	13.7	519	8	AP005621	Oryza sat
11	398	12.4	184118	8	AP005621	Oryza sat
12	328.5	10.2	150597	2	AP005005	Oryza sat
13	328.5	10.2	156603	8	AP004752	Oryza sat
14	166	5.2	3429	10	AB029919	Mus muscu
15	166	5.2	5837	6	AR447512	Sequence
16	166	5.2	7764	6	CO414477	Sequence
17	166	5.2	7764	6	AX828387	Sequence
18	166	5.2	7764	6	AF006514	Homo sapi
19	163	5.1	110000	2	PFMAL8P1_11	Continuation (12 o
20	162.5	5.1	4769	6	CQ843036	Sequence
21	162.5	5.1	4769	9	AK125995	Homo sapi
22	161	5.0	5175	6	BD079934	Cancer-as
23	161	5.0	5175	9	HUMCLIP	Human cytop
24	159	4.9	5857	6	AR274942	Sequence
25	159	4.9	5857	6	AX329796	Sequence
26	159	4.9	5857	9	HSRESTIN	H. sapiens m
27	159	4.9	5858	6	A21860	H. sapiens M
28	158	4.9	5061	10	BC076567	Mus muscu
29	157.5	4.9	7396	6	AX686226	Sequence
30	157.5	4.9	95679	8	ATT29A15	Arabidops
31	157.5	4.9	200001	8	ATCHRIV67	Arabidops
32	156.5	4.9	4190	10	BC046638	Mus muscu
33	156.5	4.9	5946	6	CQ730200	Sequence
34	156.5	4.9	5998	9	CR456526	Homo sapi
35	156.5	4.9	7318	10	AB026497	Mus muscu
36	156.5	4.9	7450	9	AB191263	Homo sapi
37	156	4.9	5176	6	CQ020760	Sequence
38	155.5	4.8	38570	1	AC000108	Helicobac
39	154.5	4.8	6965	9	HSU31906	Homo sapien
40	154.5	4.8	7636	6	AX933989	Sequence
41	154.5	4.8	7695	6	AX336423	Sequence
42	154.5	4.8	7695	6	AX933991	Sequence
43	154.5	4.8	7695	9	HSU41740	Human trans
44	154.5	4.8	7743	6	AX933993	Sequence
45	154.5	4.8	7761	6	AX933995	Sequence

ALIGNMENTS

RESULT 1
AF213627
LOCUS AF213627 Arabidopsis thaliana MOM mRNA, complete cds.
DEFINITION AF213627 6554 bp mRNA linear PLN 01-JUN-2000
ACCESSION AF213627
VERSION AF213627.1 GI:8132767
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 6554)
AUTHORS Amedeo,P., Habu,Y., Afsar,K., Scheid,O.M. and Paszkowski,J.
TITLE Disruption of the plant gene MOM releases transcriptional silencing
of methylated genes
JOURNAL Nature 405 (6783), 203-206 (2000)
MEDLINE 20279299
PUBMED 10821279
REFERENCE 2 (bases 1 to 6554)
AUTHORS Amedeo,P., Habu,Y., Afsar,K., Mittelsten Scheid,O. and
Paszkowski,J.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box
2543, Basel 4002, Switzerland
FEATURES
Location/Qualifiers
source
1..6554
/organism="Arabidopsis thaliana"

```
/mol_type="mRNA"
/cultivar="Zhuierich"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATEAT1 and phyA"
310..6315
/function="involved in the maintenance of transcriptional
gene silencing"
/codon_start=1
/product="MOM"
/protein_id="AAF7380.1"
/db_xref="GI:8132768"
/translation="MKWDEKIGLGRITVTRSLAASIPASVEQETPGLRRSRGTPST
KVTIPASATRKSERLAPSPASVSKSGIVKNSPSSLRSRNRKTEVSLQSSKSDN
SVRGSDPDIIPAGNDVSVEEDTKIPIMASARYALFRGLKESALVDASVETEKIV
VVGCSRRIPAGNDVOGKDCPPADAGSKRLPVDETSLDKGTDFPLKSVTEKIV
LDASPIVETGDDSVSGSPENLETKLDQKGTDCSPANAESKTLPVGETSLEKYPQ
KQDNTDCLPPANAESKRLPVGETSLEKDTDFPLKSTTETGKMWLVASPIVETKDD
VICSPTNLETQKLVSTGLETIVLPLKRDRTAEIELDACATVANGDDHVMSSDG
VTPSGCKNDNRPMCMCTCKRKQVNGDCNRVSCICVQVEESDNVTDQMKETGP
VTSREYENGOIQHGKSDPKFYSSVYPEYVVPVOLSDVQLQYQOTLPSKLSLSL
SKIDLGALEETLNSVRKTCDFPVMDSALKQLLTKNLELHEILDVEIKASGKLHLDK
MLTHKNGLKAVFVQATQTPTEGLLGNILEDFVQGRPKSYEHGLYSKKNALN
NENKSCQCVLLETACGQTIKLADAPILFGSLNPSHDVKVKEIKIESCERT
KIFRYSVCTVEKALILARQNMQRKQAVENRSLTHALLMWSAYLFDKLDHPSH
ETPDSGVEQIMDGVITHEFSSILSKGSENEVKLLCLLEAKHAQGTYSDDSTLFG
EDHILKSDEESPNIPWSKLLGGKPMWKYPSDTPQNRKRVOYFEGSASPKTGNGN
AKRKASDDVTDPRVTPDPPVDDDERKASGDHMGALBSPKVTILQSSCKSGTGTLL
DGNDAFGLYSMSHLSGIPEDMLASQDNGKIPDESQRHLHTVLKPMAKLCOVLHSD
ACTSMVGNFLEVIENHRIYERPATTPQAFQIALSWIAKLLVKQLLSHESLRANSE
LAFCSLDEVIYILSCMSKLFLEHTQGLQDFGTNSKQSVSTKLNVNLSLGSAT
VDEKINTKMSNDEDECKTEKRSCHVSTATRDIETKISGIIKKYKQVQKVOEH
BEKKMELNMAKDKQKLETSKSEAAVIRITCSRTSTQVGLDKLLDHNRYERKFPDIK
SEKNCLSLEQMHVAVKKLAEDACWNRKSWAALKVCVPQIQSGNKHFGSSN
ISQNPADQVICNNANVEATYADTNCKASKNVQVPEANTLGTMSGGSTQQVHEWVDV
NDEMDVSALSRQITQSQSNESHATVPEILIPADCEEPALNHLHSEDQNCDRIT
SAASDEDSVRPEVSQLENLASSPEFSLNREELVTTNERNRTHVSGFDTDNILLDQ
NREDCSLDOEIPDELAMPVQHLASVETRGAAESDQYQDICIPIPSFSLAGKQDPAA
TESNLEAEIPEQASGETVETDFAASHQGVQVTCPLSSPTGNQAPAEANIEQNTV
NTSARPHVAGPDASGVDAVIDQETMGAAQDCSLPSGSGVGTQSLDGNIEGQNTV
AQLPDGDMAIVTGGSPVSDCAQDASPMPLSPGNHPTAVNIEGLDNTSVAEPHS
GSDACEMSEISFGQVSTFANLPHEGGVEHSAGVTALVPSLLNNGTEQIAVQPVQ
IPFVFNDFLEHLEKLRSESNKKTFEKKSILKAELEKRWAVQAEFRFRKPEVE
AEHNTRTTKIENKLVIMNKLNAFLSKCTDKVSPSGAPRGRIQCLAQRAAQA
LKNYTAPOQLQASFPALVSAPIQLQOSFPAPQAPLQPAQSFPSSVRPSPALL
LNFVCPMPQRPQPLISNIAPTPTPATNGLRSPAPHLNSYRPSSTPTVATATPTS
SVPPQALITYSAVSIQQQEQPQSSLSGLSQSNNEVVCLSDDDE"
```

ORIGIN

Alignment Scores:			
Pred. No.:	1..87e-250	Length:	6554
Score:	3210.00	Matches:	620
Percent Similarity:	99.84%	Conservative:	0
Best Local Similarity:	99.84%	Mismatches:	1
Query Match:	99.81%	Indels:	0
DB:	8	Gaps:	0
SEQ3_K705_COPY_650_1270 (1-621) x APE213627 (1-6554)			
QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisAspVallyHis	20
DB	2257	CGAGCTGATGCGCTTTATCTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	2316
QY	21	ValGluLysIleLysIleGluSerCysSerGluAqTgThrLysIlePheArgLeuTyrSer	40
DB	2317	GTAGAGAAGATATAAATCGAGTCATGTTCTTGAAGACCTAAGATATCCGATTGTACTCA	2376
QY	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAenLysArgGlnAenLys	60
DB	2377	GTATGTACAGTTTGAGAAAAAGCCCTGATTCTGGCTAGGCAAAATATGCGGCAATAAA	2436
QY	61	AlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80

```

QY 441 ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMet 460
DB 3577 CGTGACGAAAGAAATTAATACGAAGTCGATCGAAATAGCTCAGAGGATGAAGAGTCATG 3636
QY 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
DB 3637 ACTGAGAGAGATGTAGCCATTATAGCAGCAACAGAGATATCGAAGACTATATAGT 3696
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 3697 GGCATAAAAGAAATACAGAAAGCAAGTCGAAAGCTTGTTACAGAGCATGAGGAAAG 3756
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer 520
DB 3757 AAAATGGAGCTGTAAATATGATATGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3816
QY 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540
DB 3817 GTGGAAGCAGCAATATCGTATTACTGTTTCAGGACCAAGTACTCAAGTGGGTGATCTC 3876
QY 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluLysSerGluLysAsnGlu 560
DB 3877 AAATGCTGGATCATTAATATGAAAGAAAGTTTGATGAATCAAAAGTCAGAAATGAA 3936
QY 561 CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGlu 580
DB 3937 TGGCTCAAAAGCTCGAGCAAAATGCAAGAGTTGCAAGAAAGAAAGTTGGCTGAGGATGAA 3996
QY 581 AlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLysLysValCysValProIle 600
DB 3997 GCGTGTGTGATTAATCGGATAAAGAGCTGGGAGCTGAAATTAAGTTGTGTTCCTCAT 4056
QY 601 GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp 620
DB 4057 CAAAGTGGCAATACAGACATTTTAGTGGTTCATCAACATTTCCCAAAATGCTCTGAT 4116
QY 621 Val 621
DB 4117 GTA 4119

```

```

RESULT 2
AX063583
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source
1. .6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310..6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
/translation="MKDKIKIGTGTIVTSLAASIPASVEQETPLRRSRGTPST
KVITPASATKSLRSPASVSKSGGIVKNSPTSLRRSRNGKTEVLSQSKSGDN
SVKRGDTPDIQRKDSVEEDTKIKPIMSARSYRALFRGLKESALVDASPNHEEL
VVGCSRRIPAGNDVQGGTDCPPPADAGSKRLPVDETSLDKGTDFPLKSVTETKIV

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,88e-250 Length: 6571
Score: 3210.00 Matches: 620
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 99.81% Indels: 0
DB: 6 Gaps: 0

SEQ3_K705_COPY_650_1270 (1-621) x AX063583 (1-6571)

QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
DB 2257 CGAGCTGATCGCTTATTCTTTTGAAGCAGCTTGAATCCATCGCATGATGATTAAGCAC 2316
QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
DB 2317 GTAGAGAAGATAAAATCGAGTCATGTTCTGAAGAAGCAATGAAGATATTCGATTTGTTACTCA 2376
QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys 60
DB 2377 GTATGTACAGTTGAGAGAAAACCCCTGATTCGGCTAGGCAAAATATGCGGCAAAATATA 2436
QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
DB 2437 GCTGTAGAGAACCTAAACCGCTCTCTCACGACGCACTGCTCATGTGGGGGGCGTCATAC 2496
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
DB 2497 TTATTGATAAACTGGATCATTTTTCACAGCAGTGAACCTCCAGATTCCAGAGTTTCATTT 2556
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly 120
DB 2557 GAACAAATCTATTATGACGGCGTGATTCATGAATTCCTGCTCCATCTTTCTTCCAAAGGT 2616
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
DB 2617 GGAGAGAAATGAAGTCAAGCTGTCTTACTTTTGGAGGCCAAGCATGCTCAGGGAAC 2676
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerSerGluGluSer 160

```

Ds	2677	TACAGCAGTGATCTACTCTATTTGGTGAAGACCATAATTAAAGTTGTGCAGATGAAGAGAGT	2736
Qy	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyAsnProMetTrpLysTrpProSer	180
Ds	2737	CCAAATATATTTGGTCAAGAGCTGTTGGGGGAAATAATCCCTATGTGGAAATACCCCTTCA	2796
Qy	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
Ds	2797	GATACTCCCCAAGGAATCGAATAACGAGTTCAGTATTTTGGAGGTTCTGAAGCGAGTCCC	2856
Qy	201	LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysLysLysLysLysLysLysLys	220
Ds	2857	AAACTGGCGATGGTGGAAATGCAAGAAGCGAAGAGAGGCTTCGTATGATGTCACTGAT	2916
Qy	221	ProArgValThrAspProValAspAspGluArgLysLysLysLysLysLysLysLysLys	240
Ds	2917	CCCCGGTCACTGATCCGCGAGTAGATGATGATGAAGAAAGAGCCCTCTGGGAAGATCAC	2976
Qy	241	MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly	260
Ds	2977	ATGGGGGCTTTGGAGTCCACAAAAGTCATAACACTCCAGTCATCATGTAATCTTCTGGT	3036
Qy	261	ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle	280
Ds	3037	CAGATGGTACATGGATGGAATGATGCTTTTGGCTTGTATTTCTATGGCGAGCCATATC	3096
Qy	281	SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer	300
Ds	3097	TCTGGATCCAGAGATATGTTAGTCTAGTCAAGATTGGGGGAAATAATACCGATGAATCA	3156
Qy	301	GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis	320
Ds	3157	CAGAGGAGCTCCACACTGTTTAAAGCCGAAGATGGCAAAACTTTGCCAGTTTTCGAT	3216
Qy	321	LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGlyTyrValIleGluAsnHis	340
Ds	3217	CTTTTCAGATGCTTGACACAGCATGGTCGGAATTTTCTCGAATATGTTAATGAAATCAC	3276
Qy	341	ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle	360
Ds	3277	CGAATCTAGAGAGCCAGCCACTACTTTTCAGGCATTCAGATAGCCCTGAGTTGGATT	3336
Qy	361	AlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer	380
Ds	3337	CGAGCCTTGTGGTAAAGCAATCTTAGCCCAAAAGATCTCTGGTCGGTCAAAATCT	3396
Qy	381	GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys	400
Ds	3397	GAATTAGCTTTCAATAGCTCTAGAGTAGAGGGGATTAATTTATTCGATATTTGCTCTGC	3456
Qy	401	MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn	420
Ds	3457	ATGAGAGTCTGTTCTCGAGCATACACAGGTTTCAGTTCGATGCTTTGGTACTAAT	3516
Qy	421	SerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal	440
Ds	3517	TCTAAACAGTCAGTGGTTAGCACAAAACCTAGTAAATGAAAGTCTCTCAGGGGCTACAGT	3576
Qy	441	ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerCysLysGluCysMet	460
Ds	3577	CGTGCAGAAAGATTAATACGAAGTCGATGCGAAATAGCTCAGAGGATGAAGAGTGCATG	3636
Qy	461	ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
Ds	3637	ACTGAGAGAGATGATGACCATTAATAGCACAGCAGCAAGAGATATCGAAAAGCTATTAGT	3696
Qy	481	GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
Ds	3697	GGCATAAAAAAGAAATACAAAGCAAGTGTGCAAAAGCTTGTACAAGAGCATGAGGAAAAG	3756
Qy	501	LysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer	520
Ds	3757	AAAATGGAGCTGTTAAATATGATGACAGCAAGACGCAAACTTGAATAGTAAAGT	3816
Qy	521	ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu	540
Ds	3817	GTGGAAGCAGCAGTAATTCGTATTACCTGTTTCCAGCAGCAGTACTCAAGTGGGTGATCTC	3876
Qy	541	LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu	560
Ds	3877	AAACTGCTGGGATCAATATTATGAAAGAAAGTTTGAATGAATCAAAAGTGAGAAAATGAA	3936
Qy	561	CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLysLysLysLysLys	580
Ds	3937	TGCCTCAAAAGTCTGGAGCAATACCGAGGTTGCAAGAGAAGTTCGCTGAGGATGAA	3996
Qy	581	AlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCysValProIle	600
Ds	3997	GCCTGTTGGATTAATCGGATAAAGAGCTGGGCAGCTAAATTTAAAAAGTTTGTGTTCCCAT	4056
Qy	601	GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp	620
Ds	4057	CAAGTGCAATTAACAGCATTTTAGTGGTTCATCAAAACATTTCCCAAAATGCTCTGAT	4116
Qy	621	Val	621
Ds	4117	GTA	4119
RESULT 3			
LOCUS			
Arabidopsis thaliana MOM gene, complete cds.			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Arabidopsis thaliana (thale cress)			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1. 10329			
/organism="Arabidopsis thaliana"			
/mol_type="genomic DNA"			
/cultivar="Columbia"			
/db_xref="taxon:3702"			
/chromosome="1"			
/map="between ATEAT1 and phyA"			
Join(705..1008,1296..2550,2674..2752,2868..3113,			
3507..3680,3974..4895,4976..5217,5778..5882,6083..7480,			
7616..7777,7915..8070,8154..8318,8452..8629,8719..8918,			
9001..9211,9285..10047)			
/product="MOM"			
Join(1301..2550,2674..2752,2868..3113,3507..3680,			
3974..4895,4976..5217,5778..5882,6083..7480,7616.			
7915..8070,8154..8318,8452..8629,8719..8918,9001..9211,			
9285..9808)			
/function="involved in the maintenance of transcriptional			
gene silencing"			
/codon_start=1			
/product="MOM"			
/protein_id="AAF73381.1"			
/db_xref="GI:8132770"			

/translation="MKKNEKIGI.TGPTTVTPSI.AASTPASVEOETPCI.PPSSPCTPST

4532	Db	GGAGAGAAATGAAGTCAAGCTGTGTCTACTATTTTGGAGGCCAAGCATGCTCAGGGA	4591
141	Qy	Tyr-SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
4592	Db	TACAGCAGTGAATCTACTCTATTTGGTGAAGACCATATTAAGTTGTCTCAGATGAAGAGAGT	4651
161	Qy	ProAsnIlePheTrpSerLysLeuLeuGlyLysLeuGlyLysAsnProMetTrpLysTrpProSer	180
4652	Db	CCAAATATATTTTGGTCAAAAGCTGTGTGGGGGAAAAATCTCTATGTGGAAATACCCCTTCA	4711
181	Qy	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
4712	Db	GATACTCCCCAAAGGAATCGAAACGAGTTCAGTATTTTGGGGTTCCTGAAGCGAGTCCC	4771
201	Qy	LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp	220
4772	Db	AAACTGGCGATGGTGGAAATGCAAGAGCGAAAGAGGCTTCTGATGATGTCACCTGAT	4831
221	Qy	ProArgValThrAspProProValAspAspAspGluArgLysAlaSerGlyLysAspHis	240
4832	Db	CCCGGGTCACTGATCCGCCAGTAGATGATGATGAAGAAAGGCCCTCTGGGAAGGATCAC	4891
241	Qy	Met-----	241
4892	Db	ATGGGTAAATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATGTGATTATGCAACTA	4951
242	Qy	-----GlyAlaLeuGluSerProLysValIleThrLeuGln	253
4952	Db	CTTTGCTGACTATCTTTCCTACAGGGGCT--TTGGAGTCCACAAAGTCATACACTCCAG	5010
254	Qy	Ser-SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273
5011	Db	TCATCATGATAAATCTTCTGGTACAGATGGTACATTCGATGGAAATGATGCTTTTGGCTTG	5070
274	Qy	Tyr-SerMetClySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293
5071	Db	TATTTCTATGGCAGCCATATCTCTGAAATCCCAGAGGATATGTTAGCTAGTCAAGATTGG	5130
294	Qy	GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla	313
5131	Db	GGGAAATACCGATGAAATCACAGAGAGGCTCCACACTGTTTTTAAAGCCGAGATGGCA	5190
314	Qy	LysLeuCysGlnValLeuHisLeuSer-----	322
5191	Db	AAACTTTGCCAAGTTTTTGATCTTTC-AGTAAGTGGCCTTTTTTCACTCCCACAACTATT	5249
322	Qy	-----	322
5250	Db	TTAGCCTTGCATATGCTTATATATAGCTGATTGCAACTGTAGTTGTTACCTGATTTCCTG	5309
322	Qy	-----	322
5310	Db	TTACAGCCAATGTGAGAGTTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTA	5369
322	Qy	-----	322
5370	Db	TTTCTTATATCTGGCTTCGTTACCAATGCACCTGTTAAATAGCAACTGTGTCACAAAAC	5429
322	Qy	-----	322
5430	Db	AGTAGGTAGTTATGTGCCTCATGTCATTCTATTTGTTTGAAGCAAGAAATTTCTGTCT	5489
322	Qy	-----	322
5490	Db	ACTTTACATGATCCATCTGTGGGAGTATATACTATATTAACCTTAGGCCTTTGTACCT	5549
322	Qy	-----	322
5550	Db	GGCTGATCAAGACATGTCAAAGTTTATCTGTTCCTGTTGGTATAGAAACTAATACAG	5609
322	Qy	-----	322
5610	Db	TGTCGTGATGCTATTTTAAAGGTAGTCTTATGTCCTTACATATTTGGCTAAATAGATGTTCCG	5669

ORIGIN

Alignment Scores:		
Pred. No.:	3, 03e-231	10329
Score:	2977.00	617
Percent Similarity:	68.51%	Conservative: 1
Best Local Similarity:	68.40%	Mismatches: 3
Query Match:	92.57%	Indels: 284
DB:	8	Gaps: 3

SEO3 K705 COPY 650 1270 (1-621) x AF213628 (1-10329)

Qy	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis	20
Db	4172	CGAGCTGATGCGTTTATTCTTTTTTGGAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	4231
Qy	21	ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40
Db	4232	GTACAGAGAAGATAAAAATCGAGTCATGTTCTCGAAGAACTAAGATATTCCGATTGTACTCA	4291
Qy	41	ValCysThrValGluGluLysAlaLeuLeuAlaLeuAlaArgGlnAsnLys	60
Db	4292	GTATGTACAGTTGAAGAAAAAGCCCTGATTCTGGCTAGGCAAAATAAGCGCAAAATAAAA	4351
Qy	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
Db	4352	GCTGTACAGAACCTTAAACCCGCTCTCTCACGCACGCACTGCTCATGTGGGGGGCGGTCAATC	4411
Qy	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
Db	4412	TTATTTGATTAACCTGGATTCATTTTTCACAGCAGTGAACTCCAGATTCCAGAGTTTCATTT	4471
Qy	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
Db	4472	GAACAATCTATTATGGACGGCGGTGATTTCATGAATTCTTCGTCCATACATTTCTTCCCAAGGT	4531
Qy	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140

QY	322	-----	322	QY	580	GlulalaCysTrrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro	599
Db	5670	CTGTCGTGCCATATACTTCTGTGATTATACACGGTCTCGGTCTATCAAAATTTGTACTAA	5729	Db	6749	GAAGCCTGTGCGATTAAATCGGATTAAGAGCTGGCGAGCTAAATTTAAAGTTTGTGTCC	6808
QY	323	-----	326	QY	600	IleGlnSerGlyAsnAsnLysHiePheSerGlySerSerAsnIleSerGlnAsnAlaPro	619
Db	5730	AAGGTATTTTGCATGTGTGATTGGTTAAACAGATTATTTTGTTCAGGATGCTTGACACA	5789	Db	6809	ATTCAAAGTGGCAATACAAAGCATTTTAGTGTTCATCAACATTTCCAAAATGCTCT	6868
QY	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHieArgIleTyrGluGluPro	346	QY	620	AepVal	621
Db	5790	AGCATGGTCGGAATTTTCTCGAATATGTATTGAAAATCACCGAATCTACGAAGAGCCA	5849	Db	6869	GATGTA	6874
QY	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu-----	357	RESULT 4			
Db	5850	GCCACTACTTTTCAGGCATTCAGATAGCCCT-GGTATGACAGCATTTACTTTGTATAATT	5908	AC026875/c			
QY	357	-----	357	LOCUS			
Db	5909	TATGCATTTGTTCCITTCATCATCTGCTTTTGTAGTAATGCTCTCAGAAAGGCAGCACTCC	5968	DEFINITION	AC026875	120965 bp	DNA linear
QY	357	-----	357	AC026875			Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
Db	5969	TTTAGTTTAACTTTCCAAATCATAGGATTCAAATATCCAAATTAACCTGGCCCTTTGATCGCTG	6028	I, complete sequence.			
QY	358	-----	359	AC026875			
Db	6029	CATAATATATGAATAGTTGCATACACTGATACGTTGTTTAATGATGCAATTTTCAGAGTTGG	6088	VERSION	AC026875.4	GI:7651711	
QY	360	IleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerHisValArgAlaAsn	379	KEYWORDS			
Db	6089	ATTGACAGCCTGTGTGTAAGCAAAATCTTAGCCACAAAGAATCTCTGTCGCTGCAAAAT	6148	SOURCE			
QY	380	SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer	399	ORGANISM			
Db	6149	TCTGAATTAGCTTTCAATAGCTCTAGAGTAGAGGTGGATATATTTATTCGATATTTGCTC	6208	Arabidopsis thaliana (thale cress)			
QY	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419	Arabidopsis thaliana			
Db	6209	TGCTGAAGAGCTGTCTCTGGAGCATACACAAGGTTTCAGTTCGATTTGCTTTGCTACT	6268	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
QY	420	AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr	439	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Db	6269	AATTTCTAAACAGTCAGTGTGTAGCACAAAACACTAGTAAATGAAAGTCTCTCAGGGGTACA	6328	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.			
QY	440	ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys	459	1 (bases 1 to 120965)			
Db	6329	GTGCGTGACGAAAGATTATACGAAGTCTGATGCGAAATAGCTCAGAGGATGAGAGTGC	6388	2 (bases 1 to 120965)			
QY	460	MetThrGluLysArgCysSerHisThrSerThrAlaThrArgAspIleGluLysThrIle	479	3 (bases 1 to 120965)			
Db	6389	ATGACTGAGAAGAGATGTAGCCATTATAGCACAGCAACAGAGATATCGAAAAGACTATT	6448	Ecker, J.R.			
QY	480	SerGlyIleLysLysLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGlu	499	Direct Submission			
Db	6449	AGTGGCATAAAAAGAAATACAGAAAGCAAGTGCAAAAGCTGTACAGAGCATGAGGAA	6508	Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,			
QY	500	LysLysMetGluLeuLeuAsnMetTyrAlaAspLysGlnLysLeuGluThrSerLys	519	Department of Biology, University of Pennsylvania, 38th Street and			
Db	6509	AGAAAATGGCTGTTAATATGTATGATGACAGACAGACAGAACTTGAACACTAGTAAA	6568	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA			
QY	520	SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp	539	Hamilton Walk, Philadelphia, PA 19104-6018, USA			
Db	6569	AGTGTGAAGCAGCAGTAATTCGTATTACCTGTTTCAGGACCACTACTCAAGTGGGTGAT	6628	On Apr 27, 2000 this sequence version replaced gi:7543638.			
QY	540	LeuLysLeuLeuAspHisThrTyrGluArgLysPheAspGluIleLysSerGluLysAsn	559	Location/Qualifiers			
Db	6629	CTCAAACTGCTGATCATATATATGAAGAAAGTTTGTATCAAAATCAAAATGAGAAAAAT	6688	1. 120965			
QY	560	GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp	579	/organism="Arabidopsis thaliana"			
Db	6689	GAATGCCTCAAAAGTCTGGAGCAATTCACGATGTTGCAAGAGAAAGTGTGCTGAGGAT	6748	/mol_type="genomic DNA"			
				/db_xref="taxon:3702"			
				/chromosome="1"			
				/clone="T6D22"			
				Join(410..559,596..758,828..865,1681..1731,1903..2253,			
				3562..3999,3096..3146,3240..3322,3493..3659,3781..3819,			
				3992..4021,4069..4246,4597..4703,4824..4952,5058..5122,			
				5239..5394,5549..5709,5812..5843,6103..6266,6382..6461,			
				6658..6723,6881..7170,7235..7264,7463..7570,7663..7760,			

7939.. 7999, 8364.. 8440, 8552.. 8623, 8702.. 8778, 8898.. .8978,
9337.. 9439, 9593.. 9751, 9883.. .10281)
/notes="similar to AT.1.24-9 gene product gb|AA07881.1;
similar to EST emb|Z35021.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.1"
/protein_id="AAF79821.1"
/db_xref="GI:8778822"
/translation="MIGHNFCORNLDSLFPTTKNSLSQTIIVAVPKISFVNRCSGT
TOVTKGKSSSPNOADEKCGKRVVAASRVCGLRLTLNKLNOQFPRD
LTVKVLITQGGIDGEMLVFDFGFDLASKFLDFCLLRMSFPHTNTSNVVPV
SQLVDSARRNLVNAISLYFVVRKLSHLVMSLIPHTHPAFSINHLATSRTS
SLRSSPPTLARSICFNSIRFLSSDMPKQKQKRDHAEQKWQKPKDPAFESGDS
SATVAEAVNNFGSLKESNTNAPLVSQTTNHRVQNLVWKPKSYTAVSGSSAT
EYKTSIAVQIGSGGQTKVGLNLSKIFGGLNLEKFSVDKSTYCHAOIRATFPKFE
KIDORTMIEMVSKGLATLEVLKSHGSLFWYAGHKGAVAKNSFGNI FIVSAQG
VILLADDLVTFMDKIPHIITVROQIYFKKHDEANAFYYSFSLMRLGELRLQK
KQNGHQRDLDDYNSKVALQKGNVDHLAAVVVAVTELGNGKQPQYSEIISPCR
WRLTNHVLFTSRKSVTFFAADFALCEGIATSVCRALDEADISVPASDHVKVQ
GBILEGLVARIYSSQSRDMENVLRDHPDPCDGNLDGLSLREICAAHNEKQOM
RALLASVGFSPCSQVWEGFDESHPSADKSVITKLOSPADYSTKLEQWMLMKE
KRLPAFKCYHNFRAEDISPNLFYKLVVHVSQSGFRYHKEWHEMPSLWPLRYGF
FVDNLFKSGARDLMALXSIDNASENDGRGEKGLADDDANLIMKFKELTYKLTFL
IRNGLSILFKGAAATKYILKQMIKWTGSDGQKELCQDLEMAIYIRKCGNDLS
SSYLSSEAPFLEQVAKRSPKNIHLIGSAGNLVREDFLAIYDGLDERGDLVKKQGV
TPATPEPAVEAVQDEGLIVFPVSYIELMDGSGIPGSAKALCKELINAPGFGDD
RPVHTLMGDLVKGWPKVADERRKQPSIMLADKNAPNEDVWROI EDMCRTRASAV
PIVADSEGTDTNPSILDALAVFMFVLOVNRHPSKLDKESNAGVYLLMFHLYSEGN
RNEFSSELERFGSLIKMFLKSDRTPDPVKSVLREGIDLFNLHRSRHRGLESYK
TYAENWKEKQRLDVLANSSELYSSIQVFFESMVHVRBEELKTIAGDYKPPSESR
KGSVIFVAINPATQVLSLEKLAANPTMSFLEKKKSIQELKERSHVTLAHRS
HGVAITVQSOHLNREVPVELTILYNDKMAALTARHVSQDGETVVSKNEMPHVTLWT
ASGVTAKEANTLPOLYLEGKASRLVIDRPVSISSGLEPF"
join(11454..11915,12009..12885,12925..13073,14200..14235,
14754..14777,14903..15370,15466..16353)
/notes="putative elongation factor 1-a
gb|AAF13128.1|AF124801.1; similar to ESTs gb|AW398653.1,
gb|Z17990.1, gb|AW220334.1, gb|AW622026.1,
dbj|AV439770.1, gb|T45275.1, and gb|AW267999.1,
gb|T04414.1, gb|AW221780.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.2"
/protein_id="AAF79822.1"
/db_xref="GI:8778823"
/translation="MGKEKPHINIVIGHVDSGKSTTTGHLIYKLGIDKRVIERFEK
EAEWMNRKFKYAWLDDKLKABERGITIDIALMKPFTTKYCTVIDAPGHRDFIKNM
ITTSQADCAVLIIDSTTGFEAGISKDQTRHALLAFTLVGKQMICCNKMDATTP
KYSKARYDRIKEVSYLKKGVNPKDIPFVPSGPEGNMIERSTNLDWKYKPTLLE
ALDOINEPKRPSDKPLRLQDQYKIGGIVPGRVETGMIKPGMVVTFAPGLTTE
VKSVMHESLLEALPGDNVGNVAVKDLKRGVANSKDDPAKGAANTSOVIL
MNHQGIINGYAPVLDCHTSHIAVKSSEILTKIDRSKGEIEKEPKFKNGDAGWVNM
TPTKPMVTFSEYFPLGRFAVDRMQTVAVGIVKSDVKDPTGAKVTKAAVKKGAK
DSVSTFIYVYLAVLCSCFVSPSERCNRVLRMRWYTHILPPHPSPLCDIV
MTEHNFVLTKGKPKPHINIVIGHVDSGKSTTTGHLIYKLGIDKRVIERFEKAAE
MNRKSFYAWLDDKLKABERGITIDIALMKPFTTKYCTVIDAPGHRDFIKNMITGT
SQADCAVLIIDSTTGFEAGISKDQTRHALLAFTLVGKQMICCNKMDATTPKYSK
ARYDEIKESVSLKGVNPKDIPFVPSGPEGNMIERSTNLDWKYKPTLLEADQ
INHRKPSDKPLRLQDQYKIGGIVPGRVETGMIKPGMVVTFAPGLTTEVKS
EMHESLLEALPGDNVGNVAVKDLKRGVANSKDDPAKGAANTSOVILMNH
GOITNGYAPVLDCHTSHIAVKSSEILTKIDRSKGEIEKEPKFKNGDAGWVNMTPK
PMVTFSEYFPLGRFAVDRMQTVAVGIVKSDVKDPTGAKVTKAAVKKGAK"
complement(join(19245..20132,20226..20687))
/notes="putative elongation factor 1-a
gb|AAF13128.1|AF124801.1; similar to ESTs gb|AW034714.1,
gb|A812185.1, emb|Z17990.1, gb|AW267999.1, gb|N96009.1,
dbj|AV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"

CDS

Alignment Scores:
Pred. No.: 120965
Score: 2977.00
Percent Similarity: 68.51%
Conservative: 1

/translation="MGKEKPHINIVIGHVDSGKSTTTGHLIYKLGIDKRVIERFEK
EAEWMNRKFKYAWLDDKLKABERGITIDIALMKPFTTKYCTVIDAPGHRDFIKNM
ITTSQADCAVLIIDSTTGFEAGISKDQTRHALLAFTLVGKQMICCNKMDATTP
KYSKARYDRIKEVSYLKKGVNPKDIPFVPSGPEGNMIERSTNLDWKYKPTLLE
ALDOINEPKRPSDKPLRLQDQYKIGGIVPGRVETGMIKPGMVVTFAPGLTTE
VKSVMHESLLEALPGDNVGNVAVKDLKRGVANSKDDPAKGAANTSOVIL
MNHQGIINGYAPVLDCHTSHIAVKSSEILTKIDRSKGEIEKEPKFKNGDAGWVNM
TPTKPMVTFSEYFPLGRFAVDRMQTVAVGIVKSDVKDPTGAKVTKAAVKKGAK"
complement(20787..20857)
/product="tRNA-Lys"
/note="codon recognized: AAA; putative tRNA-Lys"
/evidence=not experimental
complement(join(21956..22189,22535..22600,22713..22929,
23041..23126))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translation="MCLNRLPSATIPRAEBREECQSMFGRNHQLLLLSFLCKRPA
ITMNGGGGGGGGGSGPTAAAAAALOKKALLQRTDITTSVVDNFTQIVNVSRV
SDPPVKNSQETYMWRASRMVQAADSLKLKSELKQTAIFSGFASLNDHVEQRIEF
DQAEKTNRLARIADDSANKELESYHSSAQRLLDI"
join(23576..23689,24080..24258,24347..24522,24635..24726)
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translation="MTGLARLVAPMIILLFIPIELVKAETITLTPETFSKUALSNN
KALQNGICWCSFYSLSDTTRCLLSPCLLLVQKEKDTAMFVKFCVPCWCKCKLIG
NLWEDLKAMEGDDIEVEGDCGTSRAVCTKVEIHSYPTFMLFYNGEEVSKYKGRD
VESLAKFVVEETKAEAKQLEDKEL"
complement(join(25423..25563,25675..26048,26127..26330,
26585..26783,26941..27842,27936..27991,28208..28413))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translation="MDSVRRKINNTTQSSSPKSPKSFVSVNRPALAAAANSTIRP
SKSVLLFIPLISAFASFLISPMAGEKMTNALTFGKISQEAAYTVTKAWQGLVTLFC
IGAMWALSGISLHRAKPAARGETKSFSEKTDQFSLNRQLELGIKKKQDSVEF
PKSRPALPAQSLPLVPIHHQTLTGSAHSGGGGDKLNRNGSQISSFSTPSKQMS
PSMYLVPSPPSVSNRLSGQDKAVSSPWSRRSSAKDITTEBEQLELAEIDKELTE
SAGKMTPTTPVSGFAMASPTVGGSTGASGATSTPLRPVMSPAQKFTPPKKE
GDFPNMSLERAIEGFIHLGVYPOIEDWRDLRQWCSVLLKPLNKTSHIQVMT
ASKLGVNTVQVSGDLPTNGTATTALPVDRTKQWQPSYSLDEDALLHQLRANLQAI
DASMQLTENQFQQQQQQQQAALIPWQECVDAISEHRLQGLMKGWVKGLLP
RSSIPADYTVKIRALAGTCVKNTEYNGGATREKKNKWSLEPPTDSHLLIYLCFV
LEHPKMLHLDPSYTGQAQSNPLFLVLPKPEPPEKYIAVSGVSTLHPGACVL
AVDQSGPTFFALYWDKQVQFTLQGRALTALWDSMLLICHRIKVGYGGVGRMNLGSSALN
ILQVDSVSDTD"

CDS

Alignment Scores:
Pred. No.: 8,826-230
Score: 2977.00
Percent Similarity: 68.51%
Conservative: 1

CDS

Best Local Similarity: 68.40% Mismatches: 3			
Query Match: 92.57% Indels: 284			
DB:	8	Gaps:	3
SQ3_K705_COPY_650_1270 (1-621) x AC026875 (1-120965)			
QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspVallyshis	20
DB	63512	CGAGCTGATCGCTTTATTTCTTTTGAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC	63453
QY	21	ValGluLysIleLeuIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer	40
DB	63452	GTAGAGAAGATAAAATCGAGTCGTCTCTGAAGAAGAACTAAAGATATTTCGATTTGACTCA	63393
QY	41	ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys	60
DB	63392	GTATGTACAGTTGAAGAAAACCCCTGATTCTGGCTAGGCCAAAATAAGCGGCAATAAA	63333
QY	61	AlaValGluAsnLeuAsnArgSerIleuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
DB	63332	GCTGTAGAGAACCTAAACCGCTCTCTCACGCACGCACTGCTCATGTGGGGGCGTCATAC	63273
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
DB	63272	TTATTGTGATAAACTGGATCATTTTTCACAGCAGTGAAACTCCAGATTTCAGGAGTTTCATTT	63213
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
DB	63212	GNACAACTATTATTGAGCGCGTGATTCTATGAATTTCTGCTCATACTTTCTTCCAAAGGT	63153
QY	121	GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr	140
DB	63152	CGAGAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATGCTCAGGAACT	63093
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer	160
DB	63092	TACAGCAGTGATTCTACTATTGTGTGAAGACCATATTAAAGTTGTTCAGATGAAGAGGT	63033
QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer	180
DB	63032	CCAAATATATTTTGTTCAAAGCTGTGTGGGGGAAAAAATCCTATGTGGAAATACCCCTCA	62973
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
DB	62972	GATACTCCCAAGGAATCGAAAAACGAGTTCAGTATTTTGAGGGTTCGAAGCGAGTCCC	62913
QY	201	LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp	220
DB	62912	AAAACTGGCGATGGTGGAATGCCAAGAAGCGAAAGAGGCTTCGTGATGATGTCATGAT	62853
QY	221	ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis	240
DB	62852	CCCCGGTCACTGATCCGCCAGTAGATGATGATGAAGAAAGGCCCTCTGGGAAGGATCAC	62793
QY	241	Met-----	241
DB	62792	ATGGGTAAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA	62733
QY	242	-----GlyAlaLeuGluSerProLysValIleThrLeuGln	253
DB	62732	CTTTGCTGACTATCTTTCTACAGGGGCT- TTGGAGTCCAAAAAGTCTAAACACTCCAG	62674
QY	254	SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu	273
DB	62673	TCATCATGTAATCTTCTGGTACAGATGGTACATTGGATGGAAATGATGCTTTTGGCTTG	62614
QY	274	TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp	293
DB	62613	TATTTCTATGGGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGG	62554
QY	294	GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla	313
DB	62553	GGGAAAAATACCGATGAATACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGCA	62494
QY	314	LysLeuCysGlnValLeuHisLeuSer-----	322
DB	62493	AAACTTTGCCAAGTTTTTGCATCTTTC-AGTAAGTGGCCTTTTTCACCTCCACAACTTATT	62435
QY	322	-----	322
DB	62434	TTAGCCTTGCATATGCTTATATATAGCTGATTGCAACTGTAGTGTGTACCTGATTTCTCTG	62375
QY	322	-----	322
DB	62374	TTACAGCAAAATGTGAGAGTTTTTATCTTCAACTATATCCATCCGTTTAAAGCATATTTTA	62315
QY	322	-----	322
DB	62314	TTTCTTATATCTGGCTTCGTTACCCTGTTTAAATGAGCAACTGCTGCACAAAAC	62255
QY	322	-----	322
DB	62254	AGTAGGTAGTTATGTGCTCATGTCAATTCATTTGTTTATTTGAAGCAAAAGAAATTTCTGTCT	62195
QY	322	-----	322
DB	62194	ACTTTACATGATCCATCTGTGGAGTATATAACTATATATAAACCCTTAGGCCCTTTGTACCT	62135
QY	322	-----	322
DB	62134	GGCTGATCAAGACATGTCAAAAAGTTTATCTGTTGCTGTTGGTATAGAACTAATAACAG	62075
QY	322	-----	322
DB	62074	TGTCGTGCTATTTTAAAGGTAGTCTTATGCTTCCACATATTGGCTAATAGATGTTTCGG	62015
QY	322	-----	322
DB	62014	CTGTCGTGCCATATACCTTCTGTGATTATCACGGTGCTCCGCTCATCAAAATTTGACTAA	61955
QY	323	-----AspAlaCysThr	326
DB	61954	AAGTATTTTGCATGTGTGATTGTTAAACAGATTTATTTGTTTTCAGGATGCTTGCACA	61895
QY	327	SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro	346
DB	61894	AGCATGGCGGAAATTTTCTCGAATATGTTATTGAAAATCACCGAATCTTACGAGAGCCA	61835
QY	347	AlaThrThrPheGlnAlaPheGlnIleAlaLeu-----	357
DB	61834	GCCACTACTTTTCAGGCATTCAGATAGCCCT-GGTATGACAGCATTTTACTTTTGATAATT	61776
QY	357	-----	357
DB	61775	TATGCATTTGTTTCCCTTCATCATCTGCTTTGTTTGTAGATGTCTCTCAGAAGCGAGCACTCC	61716
QY	357	-----	357
DB	61715	TTTAGTTTAACTTTTCCAATCATAGGATTCAAATATCCATTAATCGCCCTTTTGATCGCTG	61656
QY	358	-----SerTrp	359
DB	61655	CATAATATATGATAGTTGACATACCTGAATAGTTGTTTAATAATGCAITTTTTCAGAGTTGG	61596
QY	360	IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn	379
DB	61595	ATTGAGCGCTTGTGGTAAAGCAAAATCTTAGCCCAAAAGAAATCTCTGGTCGGTGCANAAT	61536
QY	380	SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer	399
DB	61535	TCGTAATAGCTTTCAAATGCTCTAGAGTAGAGGGTGGATTATTATTTCGATATTGTCC	61476
QY	400	CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr	419
DB	61475	TGCAATGAAGAGTCTGTTCTCTGGAGCATACACAAGGTTTGCAGTTCGATTGCTTGTGACT	61416

```

QY 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr 439
Db 61415 AATTTCTAAACAGTCAGTGGTTAGCACAAACTAGTAATGAAAGTCTCTCAGGGGTACA 61356
QY 440 ValArgAspGluLysLeuValSerMetArgAsnSerSerGluAspGluGluCys 459
Db 61355 GTGGTGACAAAGATTATACGAGTCTGATCGGAATAGCTCAGAGGATGAAGATGC 61296
QY 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspLeuGluLysThrile 479
Db 61295 ATGACTGAGAAGAGATGTAGCCATTATAGCACAGCAACAGAGATATCGAAAGACTATT 61236
QY 480 SerGlyLysLysLysLysLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
Db 61235 AGTGGCATAAAGAAATACAAAGCAAGTGCAGAAAGCTTGTACAAGAGCATAGGAA 61176
QY 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
Db 61175 AAGAAATGGAGCTGTTAATATATGATGCAAGCAAGAGCAAGCAACTTGAACTAGTAAA 61116
QY 520 SerValGluAlaAlaValLysLeuValLysLeuValLysLeuValLysLeuValLys 539
Db 61115 AGTGTGGAGCAGCAGTAATTCGTATTACCTGTTACGGACCACTACTCAAGTGGTGAT 61056
QY 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluLysSerGluLysAsn 559
Db 61055 CTCAAACTGCTGGATCATATATTAAGAAAGAGTTGATGAAATCAAAAGTGAGAAAAAT 60996
QY 560 GluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
Db 60995 GAATGCTCTAAAGTCTGGAGCAATGCACGATGTTCGCAAGAGAGTTGGCTGAGGAT 60936
QY 580 GluAlaCysTrpIleAsnArgLysSerTrpAlaAlaLysLeuLysValCysValPro 599
Db 60935 GAAGCTGTGGATTAATCGATTAAGAGCTGGCAGCTAAATTAAGTTGTGTCC 60876
QY 600 IleGlnSerGlyAsnLysHisPheSerGlySerSerAsnLysSerGlnAsnAlaPro 619
Db 60875 ATTCAAAGTGGCAATACAGCATTTTAGTGGTTTCATCAACATTTCCCAAAATGCTCCT 60816
QY 620 AspVal 621
Db 60815 GATGTA 60810

RESULT 5
AX063582
LOCUS AX063582 10329 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100801.
ACCESSION AX063582
VERSION AX063582.1 GI:12541306
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1
REFERENCE
Habu.Y., mitelsten Scheid,O., Amedeo,P. and Paszkowski,J.
Gene involved in epigenetic gene silencing
Patent: WO 0100801-A 1 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
Location/Qualifiers
1..10329
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
1009..1295
2551..2673
2753..2867
3114..3506
3681..3973
4896..4975
intron
intron
intron
intron
intron

```

```

intron 5218..5777
intron 5883..6082
intron 7481..7615
intron 7772..7914
intron 8071..8153
intron 8319..8451
intron 8630..8718
intron 8919..9000
intron 9212..9284
ORIGIN
Alignment Scores:
Pred. No.: 5,31e-231 Length: 10329
Score: 2974.00 Matches: 617
Percent Similarity: 68.40% Conservative: 0
Best Local Similarity: 68.40% Mismatches: 4
Query Match: 92.48% Indels: 284
DB: 6 Gaps: 3
SEQ3_K705_COPY_650_1270 (1-621) x AX063582 (1-10329)
QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHis 20
Db 4172 CGAGCTGATGCGTTTATTCTTTTGAAGCAGCTTGAATCCATCCGATGATGTTAGCAC 4231
QY 21 ValGluLysLysLysLysLysGluSerCysSerGluArgThrLysLysLysPheArgLeuTyrSer 40
Db 4232 GTAGAGAAGATAAAATCCAGTCACTTCTGAAAGAACTAAAGATATTCGATTTGTACTCA 4291
QY 41 ValCysThrValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
Db 4292 GTATGTACAGTTGAAAGAAAGCCCTGATCTCGGTAGGCAAAATATGCGGCAAAATAAA 4351
QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 4352 GCTGTAGAGAACTTAACCGCTCTCTCAGCAGCACTGCTCATGTGGGGGGGCTCATAC 4411
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 4412 TTATTTGATAAATCGATCATTTTTCACAGCAGTGAAACTCCAGATTCAGGAGTTTCATTT 4471
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly 120
Db 4472 GAACAACTTATTATGACCGCGGTGATTCATGAATTCCTGCTCCATATCTTTCTTCCAAAGGT 4531
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 4532 GGAAGAAAGAAATGAAGTCAAGCTGTGCTACTTTTGGAGGCCAAGCATGCTCAGGGNACT 4591
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisLysLysLeuSerAspGluGluSer 160
Db 4592 TACAGCAGTGATTTCTACTCTATTTTGTGAAGACCATATTAAGTTGTGAGATGAAGAGAGT 4651
QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 4652 CCAAAATATATTTTGGTCAAAAGCTGTGGGGGGGAAAAATCCTATGTGGAATACCCCTTCA 4711
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
Db 4712 GATACCTCCCCAAAGGAATCGAAACGAGTTAGTATTTTGGAGGTTCTGAGGCGAGTCCC 4771
QY 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysLysLysLysLysLysLysLys 220
Db 4772 AAAACTGGCGATGTTGAAATGCAAGAGCAAGAGAGCTTCTCATGATGTCATGAT 4831
QY 221 ProArgValThrAspProValAspAspGluArgLysLysLysLysLysLysLysLysLys 240
Db 4832 CCCCCGGTCACTGATCCGCCAGTAGATGATGATGAAAGAAAGGCCCTCTGGGAAGGATCAC 4891
QY 241 Met----- 241
Db 4892 ATGGGTAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTA 4951

```

QY 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
DB 4952 CTTTGTGCTACTATCTTTCTACAGGGCT-TTCGAGTCCACCAAAAGTCATAACACCTCCAG 5010
QY 254 SerSerCysLysSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu 273
DB 5011 TCATCATGTAAATCTTCTGGTACAGATGGTACATTTGGATGGAAATGATGCTTTGGCTTG 5070
QY 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp 293
DB 5071 TATTTCTATGGCAGCCATATCTCTGAAATCCAGAGAGATATGTTAGCTAGTCAAGATTGG 5130
QY 294 GlyIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
DB 5131 GGGAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTTAAAGCCGAGATGGCA 5190
QY 314 LysLeuCysGlnValLeuHisLeuSer----- 322
DB 5191 AACCTTTGCCAAGTTTTCGACTTTTC-AGTAAGTGGCCTTTTCACCTCCACAACTATT 5249
QY 322 ----- 322
DB 5250 TTAGCCTTGCAATGCTTATATATAGCTGATTGCAACTGTAGTTGTTTACCTGATTTCCCTG 5309
QY 322 ----- 322
DB 5310 TTACAGCCAAATGTGAGAGTTTTATTCTTCACTATATCCATCCGTTTAAAGCATATTTTA 5369
QY 322 ----- 322
DB 5370 TTTCATTATCTGGCTTGGTTACCAATGCNCTGTTAAATAGCAACTGCTGCACAAAC 5429
QY 322 ----- 322
DB 5430 AGTAGGTAGTTATGTGCTCATGCTCATCTTCACTTATTATTGAAGCAAAAGAAATTTCTGTCT 5489
QY 322 ----- 322
DB 5490 ACTTTACATGATCCATCTGTGGAGTATATACTATATATAAACCCTTAGGCCTTTGTACCT 5549
QY 322 ----- 322
DB 5550 GGCTGATCAAGACATGTCNAAAGTTTATCTGTTGCTGTTGGTATAGAACTAATACAG 5609
QY 322 ----- 322
DB 5610 TGTCTGATGCTATTTTAAAGTAGTCTTATGCTCTTACATATTTGGCTAATAGATGTTTCCG 5669
QY 322 ----- 322
DB 5670 CTGTGCTGCCATATCTTCTGTGATATACCGTCTCCGTCTATCAAAATTTGTACTAA 5729
QY 323 -----AspAlaCysThr 326
DB 5730 AAGGTATTTTGCNATGTGTGATTGGTTAACAGATTATTTGTTTTCAGGATGCTGCACA 5789
QY 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
DB 5790 AGCATGGTCGGAATTTTCTCGAATATGTTATTGAAATCACCAGATCTACGAAAGGCCA 5849
QY 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
DB 5850 GCCACTACTTTTCAGGCAATTCAGATAGCCCT-GGTATGACACATTTACTTTGTATAATT 5908
QY 357 ----- 357
DB 5909 TATGCAATTGTTCCCTCATCATCTGCTTGTGTTAGATGCTCTCAGAGGCGAGCACTCC 5968
QY 357 ----- 357
DB 5969 TTTAGTTTAACTTTTCCAAATCATAGGATTCAAATATCCATTAACTGGCCTTTGATCGCTG 6028
QY 358 -----SerTrp 359

DB 6029 CATATATATGAATAGTTGACATATCTGAATACGTTGTTTAATTAATGCAATTTTCAGAGTTGG 6088
QY 360 IleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
DB 6089 ATTGCAGCCTTGTGGTAAAGCAATTTCTAGCCACAAAGATCTCTGGTCGCTGCAAT 6148
QY 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
DB 6149 TCTGAATAGCTTTCAAATGCTCTAGAGTAGAGTGGATTATATTTATTCGATATTGTCC 6208
QY 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
DB 6209 TGCATGAAGAGTCTGTTCTCGAGCATACAAAGGTTTGCACTTCGATTGCTTTGTACT 6268
QY 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr 439
DB 6269 AATTTCTAAACAGTCACTGTTAGTACAAAACCTAGTAAATGAAAGTCTCTCAGGGGCTACA 6328
QY 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
DB 6329 GTGCGTGCAGAAAGATTAAATACGAAGTCGATCGAAATAGCTCAGAGGATGAAGAGTGC 6388
QY 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
DB 6389 ATGACTGAGAAGAGATGATGATTTATAGCACAGCAACAGAGATATCCAAAGACTATT 6448
QY 480 SerGlyIleLysLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
DB 6449 AGTGGCATAAAGAAAGAAATACAAAGCAAGCAAGTTCGCAAAAGCTTGTACAGAGCATGAGGAA 6508
QY 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
DB 6509 AAGAAATATGGAGCTGTAAATATGATGCAGCAAGCAAGCAAGAACTTGAAACTAGTAAA 6568
QY 520 SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
DB 6569 AGTGTGAAGCAGCAGTAACTCGTATTACCTGTTCCAGCACCAGTACTCAAGTGGGTGAT 6628
QY 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn 559
DB 6629 CTCNAACTGCTGATCATATTTATGAAGAAATTTGATGAATCAAAAGTCAGAAAAAT 6688
QY 560 GluCysLeuLysSerLeuGluMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
DB 6689 GAATGCCTCAAAGTCTGGAGCAAAATGCACGAGTTGCAAGAGAAAGTGGCTGAGGAT 6748
QY 580 GluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValPro 599
DB 6749 GAAGCTGTTGGATTAAATCGGATAAGAGCTGGCAGCTAAATTAAGTTTGTGTCTCC 6808
QY 600 IleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaPro 619
DB 6809 ATTCAAAGTGGCAATAACAAGCATTTTAGTGTTCATCAACATTTCCCAAAATGCTCCT 6868
QY 620 AspVal 621
DB 6869 GATGTA 6874

RESULT 6

ATH552638

LOCUS

DEFINITION

Arabidopsis thaliana T-DNA flanking sequence, left border, clone

340812.

ACCESSION

AJ552638.1

VERSION

left border; T-DNA flanking sequence.

KEYWORDS

Arabidopsis thaliana (thale cress)

SOURCE

Arabidopsis thaliana

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1

ATH552638 695 bp DNA linear PLN 29-MAR-2003

Arabidopsis thaliana T-DNA flanking sequence, left border, clone

340812.

AJ552638.1 GI:29368785

left border; T-DNA flanking sequence.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1

AUTHORS Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny, A.

TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites

JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)

MEDLINE 22363535

PUBMED 12445565

REFERENCE 2 (bases 1 to 695)

AUTHORS Balzergue, S.

TITLE Direct Submission

JOURNAL Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE

COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobio.gen.fr>).

FEATURES

source 1..695
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Massillauskija"
 /db_xref="taxon:3702"
 /clone="340E12"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"

misc_feature 1..695
 /notes="T-DNA flanking sequence
 left border"

ORIGIN

Alignment Scores:
 Pred. No.: 7,65e-75 Length: 695
 Score: 1033.00 Matches: 203
 Percent Similarity: 87.88% Conservative: 0
 Best Local Similarity: 87.88% Mismatches: 1
 Query Match: 32.12% Indels: 28
 DB: 8 Gaps: 1

SEQ3_K705_COPY_650_1270 (1-621) x ATH552638 (1-695)

QY 104 IleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIysGlyGluGlu 123
 |||||
 Db 2 ATTATGGACGGCGTGATTCATGAATTCCTCGTCATACCTTTCTTCCAAAGGTGGAGAGAA 61
 |||||

QY 124 AsnGluValIysLeuCysLeuLeuGluAlaIysHisAlaGlnGlyThrTySerSer 143
 |||||
 Db 62 AATGAATCAAGCTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGAATCTACAGCAGT 121
 |||||

QY 144 AspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIle 163
 |||||
 Db 122 GATTTCTACTCTATTGTGTGAAGACCATATTAAAGTTGTGCAGATGAAGAGATCCAAATATA 181
 |||||

QY 164 PheTrpSerIysLeuLeuGlyGlyLysAsnProMetTrpIysTyProSerAspThrPro 183
 |||||
 Db 182 TTTTGGTCAAAAGCTGTTGGGGGGAAGAAATCCTATGTGTGAATACCTTCAGATACCTCCC 241
 |||||

QY 184 GluArgAsnArgIysArgValGlnTyPheGluGlySerGluAlaSerProLysThrGly 203
 |||||
 Db 242 CAAGGAATCGAAACAGGTTTCAGTATTTTGAGGGTTCTGAAGCGAGTCCCAAAACCTGGC 301
 |||||

QY 204 AspGlyGlyAsnAlaLysLysArgLysLeuAlaSerAspValThrAspProArgVal 223
 |||||
 Db 302 GATGTTGGAATCGAAAGAGCGGAAGAGGCTTCGATGATGTCACGTGATCCCGGGTC 361
 |||||

QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMet----- 241
 |||||

Db 362 ACTGATCCGCCAGTAGATGATGATGAAGAAAGGCTCTGGGAAGGATCACATGGGTAAA 421
 QY 241 ----- 241

Db 422 ATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTGCAACTACTTTTGTCTGA 481
 QY 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
 Db 482 CTAATCTTTCTCAGGGGCT-TTGGAGTCACCAAAAGTCATAACACTCCAGTCATCATGT 540
 QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTySerMet 276
 Db 541 AATCTTCTGTTACAGATGGTACATTTGGATGGAAATGATGCTTTTGGCTTGTATTCTATG 600
 QY 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
 Db 601 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 660

QY 297 ProAspGluSerGlnArgGluHisThrVal 307
 |||||
 Db 661 CCGGATGAATCACAGAGGAGGCTCCACACTGTT 693

RESULT 7

AK068790 4027 bp mRNA linear PLN 24-JUL-2003

LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013163J020, full insert sequence.

ACCESSION AK068790

VERSION AK068790.1 GI:329778815

KEYWORDS FLI CDNA; CAP trapper.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS 1
 The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otono, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 4027)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hara, A., Hashizume, W., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Oeato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sakai, K., Sakai, K.,


```
Db 3995 AAAAGTTT 4003
RESULT 8
AX063609
LOCUS AX063609
DEFINITION Sequence 28 from Patent WO0100801.
ACCESSION AX063609
VERSION AX063609.1 GI:12541333
KEYWORDS
SOURCE
ORGANISM Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
AUTHORS Gene involved in epigenetic gene silencing
TITLE Patent: WO 0100801-A 28 04-JAN-2001;
JOURNAL Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source Location/Qualifiers
1..467
/organism="Brassica oleracea"
/mol_type="unassigned DNA"
/db_xref="taxon:3712"
/note="seq1-43"
ORIGIN
Alignment Scores:
Pred. No.: 2,616-31 Length: 467
Score: 494.50 Matches: 97
Percent Similarity: 84.09% Conservative: 14
Best Local Similarity: 73.48% Mismatches: 12
Query Match: 15.38% Indels: 9
DB: 6 Gaps: 2
SEQ3_K705_COPY_650_1270 (1-621) x AX063609 (1-467)
QY 111 GluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeuCysLeu 130
Db 1 GAATTCCTGTCCTACTTCTTCGGATGTTGGAGAGAAATGAGGCAAGCTGTCTA 60
QY 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu 150
Db 61 CTTTGGAGCCCAAGCATGCTCAGGGAAGTTACAGCACTGATGCTACTCTATTGGTGA 120
QY 151 AspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSerLysLeuGly 170
Db 121 GAACATGTCAGTTATCAGAT---GAAAGTCCAAATATGTTTGGTCAAGCTGTTGAGT 177
QY 171 GlyLysAsnProMetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgVal 190
Db 178 GGAAGAACCTTATGTGGAATTAATCTGTTGGATCTCTCAAGAGGAGTCGAAAAGAGTA 237
QY 191 GlnTyrPheGluGlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLys 210
Db 238 CGGCATCTTCAGGCTATGAGGAGACTACCAAGTTGGCAATGGCGAAACTTAAAGAG 297
QY 211 ArgLysLysAlaSerAspValThrAspProArgValThrAspProValAspAsp 230
Db 298 AAAAAGAGGCTTCAGATGATGTGCACA-----GTAGATAAC 333
QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGly 242
Db 334 GCTGAGAGAAAAGCCTCTGGAAGAGATCATCGGT 369
RESULT 9
AX063611
LOCUS AX063611
DEFINITION Sequence 30 from Patent WO0100801.
ACCESSION AX063611
VERSION AX063611.1 GI:12541335
KEYWORDS
SOURCE
ORGANISM Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
AUTHORS Gene involved in epigenetic gene silencing
TITLE Patent: WO 0100801-A 30 04-JAN-2001;
JOURNAL Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source Location/Qualifiers
1..515
/organism="Brassica oleracea"
/mol_type="unassigned DNA"
/db_xref="taxon:3712"
/note="2-33"
ORIGIN
Alignment Scores:
Pred. No.: 2,016-29 Length: 515
Score: 472.00 Matches: 100
Percent Similarity: 60.87% Conservative: 26
Best Local Similarity: 48.31% Mismatches: 39
Query Match: 14.68% Indels: 42
DB: 6 Gaps: 2
SEQ3_K705_COPY_650_1270 (1-621) x AX063611 (1-515)
QY 387 SerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeu 406
Db 1 TCTAGAGAGAGGTGGATGATGATATATCTTTTCTGCTACTGCAAGAGTCTATTTCGTG 60
QY 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
Db 61 GGGCGCACACAAAGGTTTCCAA----- 81
QY 427 SerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn 446
Db 81 ----- 81
QY 447 ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSer 466
Db 82 -----GAAAAGGCTGAAGATGTCATGGCTGAGAAAAGAGGTAGC 120
QY 467 HisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysTyr 486
Db 121 CATTTAGCTCAGTAAACCAGGATGTTGAAAAGACTATTAGGCACATCAAAAAGAAATGC 180
QY 487 LysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsn 506
Db 181 AGTAAGAGCCTGCATAAGCTTGTACAAACCCCTCGAGGAGAAAGATGGACCTGATGAT 240
QY 507 MetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaLysIle 526
Db 241 AGGAATGCTGTCAAGAGCAGGAACCTTCAGAAATTGTAAAAAGGTGGAAGCATCATTTAT 300
QY 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
Db 301 CGTGTCCACTATTTCAGGTATAAATCTCAGAGCTTACATGCTCTCCAAACGGCTGGAA 360
QY 545 HisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSer 564
Db 361 TGTACTTTTGAAGAAGATTGATGATCTCAAGGAGAGTTGGATGAATGCCTTGAAGT 420
QY 565 LeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
Db 421 TTAGAGCAATAAACAAGGCTGGAAGAGAAAGATTGGCTGAAGATGAAGCTGTTGGATT 480
QY 585 AsnArgIleLysSerTrpAla 591
Db 481 AGTCGGATAGAGAAATGGGCA 501
```

```
RESULT 10
AX063607/c
LOCUS           AX063607             519 bp      DNA          linear      PAT 24-JAN-2001
DEFINITION      Sequence 26 from Patent WO0100801.
ACCESSION       AX063607
VERSION         AX063607.1  GI:125411331
KEYWORDS        Brassica oleracea
SOURCE          Brassica oleracea
ORGANISM        Brassica oleracea
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
AUTHORS         Habu,Y., Mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE           Gene involved in epigenetic gene silencing
JOURNAL         Patent: WO 0100801-A 26 04-JAN-2001;
                Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
                m.b.H. (AT)
FEATURES        Location/Qualifiers
                source
                1..519
                /organism="Brassica oleracea"
                /mol_type="unassigned DNA"
                /db_xref="taxon:3712"
                /note="seq1-23"
ORIGIN
Alignment Scores:
Pred. No.:      5,58e-27      Length:      519
Score:          442.00      Matches:    89
Percent Similarity: 84.82%    Conservative: 6
Best Local Similarity: 79.46%  Mismatches: 13
Query Match:    13.74%      Indels:     4
DB:             6           Gaps:       1
SEQ3_K705_COPY_650_1270 (1-621) x AX063607 (1-519)
QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValIysHis 20
Db 324 CGCGCTGAGCTCTTTATACCTTTTGGAGCAGCCTGAATCCATCGCATGACGTTAAGCTC 265
QY 21 ValGluLysIleLysIleGluSerCysSerGluAtrGThrLysIlePheArgLeuTyrSer 40
Db 264 TTAGAGAGATTAAGCTGAGTGGTCTTGAAAGAACTAAGATTTCCGGTGTACTCA 205
QY 41 ValCysThrValGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys 60
Db 204 GTGTTTACAGTTGAAGAAAAACCCCTGATTTCTTGCT-----AGGCATAAATAG 157
QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 156 CCTCTAGAGAACCTAAACCGTCTCTCACACATGCATTGCTCATGTGGGGGGCATCATAT 97
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 96 TTATTGTAGTGAAGTGAACATCTCCATGGAGTGGNACCCAGGTTTCAGAGTTCAATTT 37
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
Db 36 GAACAGTCTATTAAAGATGCCGTAAANNACGAATTC 1
RESULT 11
AP005621/c
LOCUS           AP005621             184118 bp   DNA          linear      PLN 27-FEB-2003
DEFINITION      Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
                PAC clone:P0681G01, complete sequence.
ACCESSION       AP005621
VERSION         AP005621.2  GI:28570002
KEYWORDS        HTG.
SOURCE          Oryza sativa (japonica cultivar-group)
ORGANISM        Oryza sativa (japonica cultivar-group)
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS         Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE           Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
                clone:P0681G01
JOURNAL         Published Only in Database (2002)
REFERENCE
AUTHORS         Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE           Direct Submission
JOURNAL         Submitted (01-AUG-2002) Takuji Sasaki, National Institute of
                Agrobiological Sciences, Rice Genome Research Program; Kannondai
                2-1-2, Tsukuba, Ibaraki 305-8602, Japan
                (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsp.dna.affrc.go.jp/,
                Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT         On Feb 26, 2003 this sequence version replaced gi:22091093.
                The orientation of the sequence is from T7 to SP6 of the PAC clone.
FEATURES        Location/Qualifiers
                source
                1..184118
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="genomic DNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /chromosome="6"
                /clone="P0681G01"
ORIGIN
Alignment Scores:
Pred. No.:      6,55e-20      Length:      184118
Score:          398.00      Matches:    178
Percent Similarity: 32.47%    Conservative: 134
Best Local Similarity: 18.52%  Mismatches: 255
Query Match:    12.38%      Indels:    395
DB:             8           Gaps:       22
SEQ3_K705_COPY_650_1270 (1-621) x AP005621 (1-184118)
QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValIysHisValGlu 22
Db 19304 GATGCCATCATCATATATTGTAGTGATTTGGAATCCAAACATGACTTCGAGAGTTCTCCAG 19245
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 19244 AGGATCAGTATAGAGTCTCAATCTGAATGTGTGCTATTTTCGTTGTTATTCGTTGT 19185
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
Db 19184 ACGTGGAGGAAAAAGACTCTTATATACGGAAGCATGATCATATTCTTCAGACGAAACGTC 19125
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 19124 CAGAAATGTAATGCCCTATCGTAGGCCATTCCTGTAGTGGGGTGCATCGTTTCTCTTC 19065
QY 83 AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
Db 19064 AATAGACTTGGAGAGTTCCAAAGCATGACTACTCCAGTAAAGATTCTCGAGATGATGTT 19005
QY 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
Db 19004 CTC-----TTTCATGAATAATGTTATTTTGGAGTTTTCGAGAGTTCGACGAAATATCC 18960
QY 118 SerLysGlyGlyGluGluAsnGluValLysLeuCysLeuLeuLeuAlaLysHisAla 137
Db 18959 ACCAACGTTGAAGCCAGCAGCACTAAATGAATAATGCAATATATCTCGAGCTCAGCAGAGT 18900
QY 138 GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSer 156
Db 18899 GGATCGTTTTATTCTAGAGATATTGCTGTATATAGCAGAGAGGAGGAATATCTCAGATT 18840
QY 157 AspGluGluSerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
Db 18839 GATGTTGATTGGCCAAAGTTCTTGGACCTTCTGGTCAAAATTTACTAGGTGGGAGGCTCTCT 18780
QY 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
```

```
Db 18779 CATTGGCAATATATATCTGAGCCAGTCGCAAGAGAACCGTAGGAGATACAAAATATGGAA 18720
Qy 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysAla 214
Db 18719 GACCAATGAGAAATACCTGCAGAGAAACTGATGAGCAATATGAAACGTAGAAAATC 18660
Qy 215 SerAspAspVal---ThrAspProArgValThrAspProValAspAspGluArg 233
Db 18659 GGAGAAATCATGGATTCTCTCCGAGATCCTA-----CCTGTCAAAGACAATGAT--- 18609
Qy 234 LysAlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 18608 -----GCTGTCGCTCGAA 18594
Qy 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGly----- 267
Db 18593 AATAGTACAGCATCCAGTCTCTCATGAAACATCAGTTGATGACACTTGGCAAGAGCTAGGT 18534
Qy 268 -----AsnAspAlaPheGlyLeuTyrsrMetGlySerHisIle 280
Db 18533 ATTCCAACTGTAACAGTAATTAACAGCTCTTATGCTTTATGTCATATGCTACATTTTTT 18474
Qy 281 SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer 300
Db 18473 TATAATACACTCCATTTTGTGGGGCAGAAAGCCTTCAGGCG-----ACA 18429
Qy 301 GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis 320
Db 18428 CAGAAAGTCTTTCACACCCCACTCAACCCAGAGCTGTCAAGTGTGATGATTTGTTGAA 18369
Qy 321 LeuSer----- 322
Db 18368 TTGCGGCTACTGCCAAGTATATTAATAAATGCTTCTTTATGTTAATTTGAGAAA 18309
Qy 322 ----- 322
Db 18308 AATATTTAGTTCAGTCAAGTAGCAGACATATCTGCTCTTTTAGCAATTCAAATTAATA 18249
Qy 322 ----- 322
Db 18248 GCTAATATTTTATGTTTCAACAGCAGAGTTCACACCTTGTGCAAGAGTTATTTATCAA 18189
Qy 322 ----- 322
Db 18188 GTTGGATCTTGATAAATTTTGCTCTTCATAACGGAACAATTTCTTTGACTCTTAGTGTG 18129
Qy 322 ----- 322
Db 18128 TCATTTTGATCCATATATGTTACTTCTCCGTTTCAATGTAGACTTTCTAGCATGTC 18069
Qy 322 ----- 322
Db 18068 CCACATTCATTTAGATGTTAATGAATCCAGACATATGATGTGTGTGTAGATTCATTAACAT 18009
Qy 322 ----- 322
Db 18008 ATATATGTATGTGGCAATACTAGAAAGTCTTACATTTGCAATGTGAGGAGGATTAGTT 17949
Qy 322 ----- 322
Db 17948 GCTACTATATCCTGCTTTTAGCTCAGGTATAGAGTGTGTGATATACCTACTCCCTCGCC 17889
Qy 323 -----AspAlaCysThrSer----- 327
Db 17888 CCATAGAAAACCAACTAGTACTGATGCTGTCTAGATTCGTAGTACTAGGATGTGCA 17829
Qy 327 ----- 327
Db 17828 CATCCGGTACTAGTGTGTTTCTATGGGACGGAGGAGTGATGTTCTTACATGCTTCC 17769
Qy 327 ----- 327
Db 17768 AATATGAATTTATGACGACAGCTGTGTCCTGCAGTCTCTGTTGTTAAAAAGTTGCTCTCT 17709
```

```
Qy 327 ----- 327
Db 17708 CTTTCTGTAGAATTTATGCGAGCATGCTGTTTACCCTAAGTTTTCAGTTTAAAGCGAACTTTTCC 17649
Qy 327 ----- 327
Db 17648 TGAAGTCACCTGATCTCTAGTTTAAATATATATAGCTGATGATGCTTGGATGCTTGGAT 17589
Qy 328 -----MetValGlyAsnPheLeuGluTyrsValIle 337
Db 17588 AAAACCTTTTGCAGGAGAGCTGTCAATGTTTATGTGAGGANTTACTTGAT-TACATTTTG 17530
Qy 338 GluAsnHisArgIleTyrsGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeu 357
Db 17529 AAGAATCATCAAGTCAGTCAGGAGCCAAAGGGCATATTGTCATGCATTCACATTCACCATTTGCCCTG 17470
Qy 357 ----- 357
Db 17469 GTTAGATTATAAAATATTTTATAGGTTTTTTAGCTAAATTTTCATGGCTTAAATGCTATT 17410
Qy 358 -----SerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSe 370
Db 17409 ATTTTTTCCATGCTGCTTTTCAGTGTGGCGTGTGCTTCTCTCTTAAAGCATTAAGATCAA 17350
Qy 370 rHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGI 390
Db 17349 TCGAAGAGAGTCACTTGGCCCTTCTGTGAGAAACTTGAACCTATGAGTGCATGAGTGTCT 17290
Qy 390 uValAspTrpIleTyrsrIleLeuSerCysMetLysSerLeuPheLeuGluHisThr-- 409
Db 17289 CGCTGAATATGTTTATGAGAAGCTTAAGGATCTCTCAAGAAAAAGTTTTCGCGTAGACCAAG 17230
Qy 410 -----GlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerVa 425
Db 17229 TGAACAACGACGACGACCAATCTACTCCAGTAAATATACACATCATCTTACAAACACA 17170
Qy 425 lValSerThrLysLeuValAsnGluSer-----LeuSerGlyAlaThrValArgAspGI 443
Db 17169 GACCTCACCGAAATTCGAGAAGTACGCGATCAATTTGTCATCAGGTAAACGACAATTCATGG 17110
Qy 443 uLysIleAsnThrLysSerMetArgAsnSerSerGluAsp-----GluGluCysMe 460
Db 17109 TGATTTGGAAAAATGTTTTCATCATGAGAGGCTCCACATGATATCTTCACCGCAGAGATGAT 17050
Qy 460 tThrGluLysArgCysSerHisTyrsrThrAlaThrArgAspIleGluLysThrIleSe 480
Db 17049 ATTAGAACAGAG-----GAATTGATATC 17026
Qy 480 rGlyIleLysLysLysTyrsLysLysGlnVal-----GlnLysLeuValGlnGluHisGI 498
Db 17025 TGTTCTAGAAACTCATAGAGAGAACATGTTTTAAGGGATGAATCTTCTTGAAGAATCAC 16966
Qy 498 uGluLysLysMetGluLeuLeuAsnMetTyrsAlaAspLysLysGlnLys----- 514
Db 16965 AGAGAAAGAATTAATTAATTAACATGTTTCTCTCTTAAAGAAAGAANAATATTTCAGGA 16906
Qy 515 -----LeuGluThrSerLysSerValGluAlaAlaValIle 526
Db 16905 TAAACAAGGAACGACGACCACTGTTAGACATGTCACAGCAAGCAAGAGAGTGCAGAGCT 16846
Qy 526 eArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAs 546
Db 16845 GCAAGAAACATGCAAT-----TTGCTCGTGCACATCT 16813
Qy 546 nTyrsGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGI 566
Db 16812 TCGTAAAGTTCACATTTGAT-----TCAGAGGACAGGAGTCCCGCTGAAGTAAATAT 16759
Qy 566 uGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnAr 586
Db 16758 TGAATGTTTACTTCTCTGATGATGATTTTGAATTCATGAGATGCCAGCATTAACAA 16699
```



```
Db 138685 TTATCCTATTACATGGTGAATGACTATATCAATATATTTTGAAGAAATACATACAGTTT 138744
Qy 277 ---GlySerHisLeuSerGlyIleProGluAspMetLeuAlaSerGlnAspTTP----- 293
Db 138745 AATGGAGCGAAATATAGCAGCA-----CAAGTTCTCATTCGA-AGTAGTGAGTGGATGTTG 138797
Qy 293 ----- 293
Db 138798 TTGTTTCTACTAACTATCATTTTCCCCAAAGGTGGGAAATAAATAGTGTGGAACTCT 138857
Qy 293 ----- 293
Db 138858 TTCTTTGATCTTTACTAAATAATTTGAAGTTTCATGATGGTGCACGTACACTCATCCAGCT 138917
Qy 293 ----- 293
Db 138918 TAAGAGTAGCATCAATCATATAAAGAGTAGCGCTATCAACCTAGATTAAAGCAGCGATT 138977
Qy 293 ----- 293
Db 138978 CTCACAACTAAATCCTTAAGATATTCAACCATGAAAAATTTGTAGAGGTGCAGTTATTTC 139037
Qy 294 -----GlyLeuIleProAspGluSerGlnArgArg 303
Db 139038 ATGGACTTAGCTCATTATGCTATAGTAGTAACAAAAAGCCCTCAAAACTCAGTCAAGAGA 139097
Qy 304 Leu----- 304
Db 139098 TTACGGAGGCAAAAGTCCACTGTAAAGGAGCTCTAACTATTATTAACCTTGAATCATGAGA 139157
Qy 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis----- 320
Db 139158 CATTAGGGGAAAAAACCTCGAAGTAGAAAAAGCTGGCTGT-GTCTGCACACTCCAGCTGTCCG 139216
Qy 320 ----- 320
Db 139217 CAGTTCAGTCACACTTGTCTTAAATGGGAAAAATAATTTCTGTGAAGTCTTTTGTGG 139276
Qy 320 ----- 320
Db 139277 TTTTATCATATAACTCTTACCTTTGTGTAAAGCTGATCTCATTTATTGTAATTTTCTGTT 139336
Qy 321 -----LeuSerAspAlaCys 325
Db 139337 GCTTTGGCTTGTATCACCACCTAATGCTTTTCGCAATTAATTAATTTGCAGGACAATGTT 139396
Qy 326 ThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGlu 345
Db 139397 ACGTTCGTGGTTCAGCAGAGTCTTTGAATATCTTTTGAATTAATCATGTAGTTGTTTCAGGAG 139456
Qy 346 ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla----- 361
Db 139457 CCAAGTACATATTTTCATGCGTTGACATAGCTTTTGGTAGAATAACCCCAATATC 139516
Qy 362 -----Ala-LeuLeuValLysGlnIle 368
Db 139517 GTCGTTCTTTTAAATTTTCAGATTCAATATGACCAAAAATGCTAATGCTCGTAAATACTTT 139576
Qy 368 eLeu-----SerHisLysGluSerLe 375
Db 139577 CCTTCAGTGTGGCGAGTTCCTCTATTTCATACTTTTAAGGTGGATTAAGAGTCACT 139636
Qy 375 uValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTy 395
Db 139637 TGGCTTCTGAAAAAGGTTAAATACGAGTGCATGACGAGCTGCAGAGTGGCAAGTTGTCTTA 139696
Qy 395 rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs 415
Db 139697 TGACAGCCTA-----AAGAGAAATTTTCCAAAGAAAGCAGGTGCACTGGCAGCAA 139747
Qy 415 pCysPheGlyThrAsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLe 435
Db 139748 TTGTCAATCAACA-----TCAGTAGAAAAAGCACCACAACTTTCACAGCAGGAGAC 139795
```

```
Qy 435 uSerGlyAlaThrValArgAspGluLysIleAsnThrLys----- 448
Db 139796 CTCCTAAT-ATCTTGAGAAATGACCACATTTTCCCTAAGCAGAGAAATGGATCTTCATGA 139852
Qy 449 -----SerMetArgAsnSerSerG1 455
Db 139853 TTAATTCATGAATGGTGCACCTTCAAGAAAGGCTCATTTGTTGTCGCAAAATGGTTTCAGA 139912
Qy 455 uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyrSerThrAl 471
Db 139913 AGAGCAGGAAGTATGATGATGTTTCCAGGAACCTCACATGGAATGC--CATTTTCAACTGA 139969
Qy 471 aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysLysTyrLysLysG1 489
Db 139970 TGAACCTCTCTGCATTTGTGGAGAAAGAAATCAACTTAATTAATGACAATGTTTCTCTGTAAG 140029
Qy 489 nValGlnLysLeuValGlnGluHisGluGluLysValMetGluLeuLeuAsnMetTyrAl 509
Db 140030 AGAATACCTATCTTTTGATAAACAACAGTCTCAGATTTCAGAGTTAGAGAAATACACACA 140089
Qy 509 aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleTh 529
Db 140090 AATAAATACTGCCAGACTGAAAACA-----GT 140116
Qy 529 rCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGluAr 549
Db 140117 GTGCAAT-----TTAGTTCTGGAACATATTTTGTAGAAG 140149
Qy 549 gLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMethI 569
Db 140150 TCATGCTGAT-----GTAGAGACCCGAAATGACACAAATAAACAACACTGTTCCAGTGGTT 140203
Qy 569 sGluValAlaLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSe 589
Db 140204 CACTATGCTGATGATGCTTCTGAGGCACATGAGACTCCAGCAGCAACAACTTGAGTC 140263
Qy 589 r-----TrpAla-----AlaLysLeuLysValCysValProIl 600
Db 140264 ACTGCAATCTAATACATGCGCTGAAGAGCGACAGCTGAAGGAAAAAACTTTGTTTGAAGC 140323
Qy 600 eGlnSerGly----- 603
Db 140324 AAATCAGGCCAATTAGATCAACACTTTGATCAGCAGATTGCTTACCAGATTCAAACTT 140383
Qy 604 -----AsnAsnLysHisPheSerGlySe 611
Db 140384 TGTATGCAAGATTCATCCATTTGAGGACAAAGTAGCAATTCCTCATGCTCAGTGAAG 140443
Qy 611 rSer 612
Db 140444 CGCT 140447
RESULT 13
AP004752/c 156603 bp DNA linear PLN 21-JAN-2004
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
PAC clone:P0501G04.
ACCESSION
AP004752
VERSION
AP004752.3 GI:41052917
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS
Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, PAC
clone:P0501G04
JOURNAL
Published Only in Database (2002)
REFERENCE
2 (bases 1 to 156603)
AUTHORS
Sasaki,T., Matsumoto,T. and Yamamoto,K.
```

TITLE
JOURNAL
Direct Submission
Submitted (13-FEB-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jan 21, 2004 this sequence version replaced gi:34740248.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM
(http://www.tigr.org/cdb/glimmer/glm_form.html), RiceHMM
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.lastate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/docs/sim4.html), gap2
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI Nonredundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequences database at RGP or DBJ. Protein homologies of the coding
regions were searched against NCBI Nonredundant Protein database
with BLAST. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
This sequence of P0501G04 clone has an overlap with P0017C12 (DBJ:
AP004769) clone at 5' end and with OJ1399 H05 (DBJ: AP004090)
clone at 3' end. Detailed information on overlap and assembly
quality together with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES
source

1. .156603
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="P0501G04"
complement(join(2875..3454,3805..3815))
/gene="P0501G04.1"
complement(join(2875..3454,3805..3815))
/gene="P0501G04.1"
/note="gag-pol-like protein"
complement(4646..7455)
/gene="P0501G04.2"
complement(join(<4646..4662,5391..5429,6120..6360,
7423..7455))
/gene="P0501G04.2"
/note="start and end point are not identified"
complement(join(4646..4662,5391..5429,6120..6360,
7423..7455))
/gene="P0501G04.2"
/note="predicted by RiceHMM etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07829.1"
/db_xref="GI:41052918"
/translation="MKRLYSHKSKIGSDVVKENVPKPIERSFLDMLDSSVSGADLTQ
DKNKNINELLIQHFGDENCIDBRANKLLVDVFLLSNSKPIVPTNTTVDGIMKKLS
KDYVYSA"
10034..10234
/gene="P0501G04.3"

misc_feature

10034..10234
/gene="P0501G04.3"
/note="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
complement(13016..16261)
/gene="P0501G04.4"
complement(join(13016..13562,14371..15279,15556..16261))
/note="supported by full-length cDNA(s): AK069014"
/gene="P0501G04.4"
complement(join(13328..13562,14371..15279,15556..16211))
/note="contains EST(s): AU031908(R2510), D28321(R2510)
contains full-length cDNA(s): AK069014"
/codon_start=1
/product="kinesin light chain-like protein"
/protein_id="BAD07830.1"
/db_xref="GI:41052919"
/translation="MPGLAADNSPPVVAAPPRRLSSPLPRSRAPSPSPSTSSRAKP
RKVAQPETDEALNDPDLGFLKQARDNVGEGGAARALPEFAEAAALERRGE
GAELAMSLHVAALHCGLRHADAIVLERAVAVVTPPPPPPEGESEAEQEPPE
DOQGEWALAAFSGMWOLGTHAMLRMDESACYGKGLBIQMAALGERPRAVETC
RYLAEHVQALQFDEAEKLCRKALEIHREHSAPASLEASDRRLMALTLDAKGDYDGA
LEHLVLSMTWANGRDIEVATIDVAIGNTYLALARPDEAVFSQKALTILKSARGDD
HPSVASVVRADLYHRTGKRESKSYCENALRVTAKEAPGAAPDEAVAGGLMEIAIY
EALGDLBALKLQALKLEDSFGQSTVAGIEAQMGVLYMVGRYADSRNSFENAV
AKLRASGRKSAFFGVLLNQMLACVQLFKIDEAQLFEARAVLEQECGASHPTLG
VYSNLAAIYDAMGRVEDAIELEHVLKVRBEKLTANPDVEDEKLRLAELLKEAGRSR
NRKQKSLNLFVTSQRVKDAGRWSNFGFRS"
18235..22976
/gene="P0501G04.5"
join(<18235..18636,19319..19528,19610..20071,20165..20275,
20362..20541,21136..21420,22634..22754,22855..22976)
/gene="P0501G04.5"
/note="start and end point are not identified"
join(18235..18636,19319..19528,19610..20071,20165..20275,
20362..20541,21136..21420,22634..22754,22855..22976)
/gene="P0501G04.5"
/note="similar to Arabidopsis thaliana chromosome 1,
F1K23.23"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07831.1"
/db_xref="GI:41052920"
/translation="MPSPLLLHHHHHRLSPRFAPSPSPRTLSLHPAPRLRLRL
RASARGGGSTRDHDGARSIPICRYEALRLDLGCTARREQAVAAAAADGGAA
EALHAAVADAMVEAFIPGPDGARRSAMSTRILIQANVETKASKLKDGLGTFESEN
EPDAESVLAMAFKQVMDQLTNFRLEVPESGSEIDLDLSKPKVPADFSIRSDKEL
LAALAEAFSCVIEDARNYLGCTGGLFHKSKNSCLDSSVCIDIIISSEVYNSARR
LDSFDLVQSSHVAGKAGWPAKPSERLVKIGGDFMLWASEFVYTYKQJIDAKAFK
NTKLGHHVLANNGEVLLSHAQVVELANILDMYDPDQTLPGKTFYFNWNSPESKIK
KNNGYLNLFALLAGSSIVFLVGAIAQLCWPQSLDKRLATVSSCVSSSQSCSDHS
LDSSELOAYCVSVVEKIDGFCQDLMVDANIAGWVGLPEFFKGINCDSDHDDYDI
QSTGAISOGEOPSLVSSPIKMSHLEQNDTQETLQNTASFOQVMSRKGKVGFOPTN
RLAVNHWTNPLTKLLYEGRKISPAFLERISRPAKVPVPELLMSVNPSPFALVR
PVQDPC"
complement(23753..26108)
/gene="P0501G04.6"
complement(join(23753..24276,24367..24619,24688..24918,
25218..25348,25506..26108))
/gene="P0501G04.6"
/note="supported by full-length cDNA(s): AK106946"
complement(join(24086..24276,24367..24619,24688..24918,
25218..25348,25506..26061))
/gene="P0501G04.6"
/note="contains full-length cDNA(s): AK106946"
/codon_start=1
/product="putative GDSL-motif lipase/hydrolase protein"
/protein_id="BAD07832.1"
/db_xref="GI:41052921"
/translation="MRRLAVAAAALLLVAMHFLSLITMAAQLSPQPQDDQDDPPMP
GLPVPSPSGYSDSPPSQEPDTPPEPAPTPPQQQQPQSPLPAPRPPAP
PRTVVPQEPGWSVPPPPARVINYYTTTGCTTMLVFGDSTVDPGNRNRLQTAMKANFL

gene

10034..10234
/gene="P0501G04.3"

gene

10034..10234
/gene="P0501G04.4"

gene

10034..10234
/gene="P0501G04.5"

gene

10034..10234
/gene="P0501G04.6"

QY	294	-----GlyLysIleProAspGluSerGlnArgArg	303	QY	509	aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleTh	529
Db	57773	ATGGACTTAGCTCAGCTTATAGTGAACAAAGGCCCTCAAAACTCAGTCAGAGA	57714	Db	56721	AAATAAAATGCAGACTGAAACA	56695
QY	304	Leu-----	304	QY	529	cYSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyxGluAr	549
Db	57713	TTACGGAGGCAAAAGTCCACTGTAAAGGAGCTCTAACTATTAAACACTTTGAATCATGAGA	57654	Db	56694	GTGCAAT-----TTAGTCTGGAAACATATTTTGTAGAG	56662
QY	305	HisThrValLeuLysProLysMetAlaLysLeuGlnValLeuHis-----	320	QY	549	giysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHi	569
Db	57653	CATTAGGGGAAAAACCTGAAGTAGAAAAGCTGGTGT-GTCTGACACTCCAGCTGTCG	57595	Db	56661	TCATCTCTGAT-----GTAGACCCGGAATGACACAATAAACTGTTTCACTGGTT	56608
QY	320	-----	320	QY	569	sGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIleAsnArgIleLysSe	589
Db	57594	CAGTTCACTGACACTTGTCTCTTAATGGGAAAAATAATCTTGTTAAGTCTTTTTTTGG	57535	Db	56607	CACTATGCTGATGTATGATTTCTGGAGCATGAGACTCCAGACAGCAAACTTGAGTC	56548
QY	320	-----	320	QY	589	r-----TTPala-----AlaLysLeuLysValCysValProI	600
Db	57534	TTTTATCATATAACTCTTACCTTTGTTAAACGGCTGATCTCATTTATTTGTAATTTTCTGTT	57475	Db	56547	ACTGCAATCTAATACATGGCTGAAAGCGACAGCTGAAGGAAACCTTTGTTGGAAGC	56488
QY	321	-----LeuSerAspAlaCys	325	QY	600	eGlnSerGly-----	603
Db	57474	GCTTTGGCTTGTATCACCACCTAATAGCTTTGCATTAATTAATCTATTGCAGGACAAATGTT	57415	Db	56487	AAATCAGGCCAATTAGATCACACCTTTGATCAGCAGATGCTCTACCAGATTCAAAC	56428
QY	326	ThrSerMetValGlyAsnPheLeuGluTrpValIleGluAsnHisArgIleTyxGluGlu	345	QY	604	-----AsnLysHisPheSerGlySe	611
Db	57414	ACGTTGCTGTTGAGCAGTCTTTTGAATATCTTTTGAATATCATGCTAGTTGTTCCAGGAG	57355	Db	56427	TGTTATGCAAGAATTCATCCATTTTGAAGGAACAAAGTAGCAATTCCTCATGTCACTGGAAG	56368
QY	346	ProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAla-----	361	QY	611	rSer	612
Db	57354	CCAAAGTACATATTTTCATGCTTGAACATAGCTTTTGGTACGAATAACCAACCCCAATATC	57295	Db	56367	CGCT	56364
QY	362	-----Ala-LeuLeuValLysGlnI	368	RESULT 14			
Db	57294	GTCGGTCTTTTAATTTTCAGATTCAATATTGACCAAAAATGCTAATGCTCGTAATACTTT	57235	AB029919	AB029919	3429 bp	mRNA linear ROD 17-JUL-1999
QY	368	eLeu-----SerHisLysGluSerLe	375	LOCUS			
Db	57234	CCTTCAGCTTGGCAGTGGCCCTCTATTCTAATCTTTAAGGTGATCATATAAGAGTCACT	57175	DEFINITION			Mus musculus STAP mRNA for sperm tail associated protein, complete cds.
QY	375	uValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyxIleTy	395	ACCESSION	AB029919		
Db	57174	TGCCCTTGCTGAAAAAGGTAAAAATACGAGTGCATAGGAGCTGGCAAGGTGCTCTA	57115	VERSION	AB029919.1	GI:5525094	
QY	395	rSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAs	415	KEYWORDS	sperm tail associated protein.		
Db	57114	TGACAGCCTA-----AAGAGAAAATTTCAAAGAAAGCAGGTGGCACTGGCGAGCAA	57064	SOURCE	Mus musculus (house mouse)		
QY	415	pCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLe	435	ORGANISM	Mus musculus		
Db	57063	TTGTCAATCAACA-----TCAGTAGAAAAGACCAACCTTCACAGCAGGAGAC	57016	REFERENCE	1 (bases 1 to 3429)		
QY	435	uSerGlyAlaThrValArgAspGluLysIleAsnThrLys-----	448	AUTHORS	Ohuchi,J., Arai,T., Kon,Y. and Watanabe,T.		
Db	57015	CTCTAAT--ATCTTGAGAAATGACCACATTTTCCCTAAGCAGAGAGATGGATCTTCATGA	56959	TITLE	Sperm tail associated protein		
QY	449	-----SerMetArgAsnSerSerG	455	JOURNAL	Published Only in Database (1999)		
Db	56958	TAATTTTCATGAATGGTGCACTTCAAGAAGGCTCAATTTGTTGGCAAAATGGTTTTCAGA	56899	REFERENCE	2 (bases 1 to 3429)		
QY	455	uAspGluGluCysMet-----ThrGluLysArgCysSerHisTyxSerThrAl	471	AUTHORS	Ohuchi,J., Arai,T., Kon,Y. and Watanabe,T.		
Db	56898	AGAGCAGGAACATGATGCTTCCAGGAACCTCACATGGAATGC---CAITTTTCAACTGA	56842	TITLE	Submitted (12-JUL-1999) Tomomasa Watanabe, Laboratory of		
QY	471	aThrArgAsp-----IleGluLysThrIleSerGlyIleLysLysLysGlu	489	JOURNAL	Experimental Animal Science, Graduate School of Veterinary Medicine		
Db	56841	TGAACCTCTCGACATTTGTGGAGAAAAGAAATCACTTAATTTGACAATGTTTCTCTGTAAG	56782		Hokkaido University, Kita 18 nishi 9 kita ku, Sapporo, Hokkaido		
QY	489	nValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyxAl	509		060-0818, Japan (E-mail:watanabe@vetmed.hokudai.ac.jp,		
Db	56781	AGAAATACCGTATCTTTGATATAACAACAGCTCTCAGATTTTCAGATTAGAGAAATACACACA	56722		Tel:81-11-706-5106(ex.5106), Fax:81-11-717-7569)		
				FEATURES	Location/Qualifiers		
				source	1. 3429		
					/organism="Mus musculus"		
					/mol_type="mRNA"		
					/db_xref="taxon:10090"		
					/chromosome="8"		
					/map="51-53"		
					/sex="male"		
					/tissue_type="testis"		
					/dev_stages="adult"		
					/note="D8Mit198(51) - D8Mit138(53)"		
				gene	1. 3429		
					/gene="STAP"		
					92. 3160		
					/gene="STAP"		
					/codon_start=1		
					/product="sperm tail associated protein"		


```

/protein_id="BAA82514.1"
/db_xref="GI:5525095"
/translation="MLKLGELRTAKGLKNEAGERDRVSNLNLKLSQLDILKNLHD
VCRQKQTLQENGLCVVEEAMLKANHKKQAQTLVFTDQMDFRVKNQYHLRQLQQLKK
KILTLQOELFRFTQELQASYSLLQYQSILEKOTSLLVLRHCKLKEDEVILYEEM
GNHSGTEGKHLAOLAGAKILSLERSLNLYRDKYOTSLSNTELLSECVKMLEG
ELSLGLQDPENKDPKVIYITSPCVIQEHQETLKLSEVWKQVSEQDDLIQELRNK
LACNSLIVREBALIKLQAEFASYTATHRHPTSSDCEDDIITKILKHQEQKDSQCL
HVBEYQNLVLDLMELEAVSEQKKMKMMKMLDLHLGLRETSYVIEKKOKETVFL
QYRLQLOQYTSQKLSKDKLQDKODERLEHLEKLMQVQNSLREKEAELEKLCQ
TKKELSTLEQARQSKIDCEALRAEIOKLDKLEAREOLKVSQNLQTSKEEAHL
SASSLEDAHKKIENCLLODKQKEEVIKQLOSLKHLKQKSSKIEERKKNRQRLQSL
SELSEQRRLSNAEKSLLOKTLDEDEKKIDELFHSQVSEQKQELTNSIRKLEEE
LLEIKGLERKQKLSKEQEALEUEIEALQKRAKKEKMAKHLKLDDEKENIQ
AEITSSHLLSSLNKYSQKVIQELNABEIAKODSIMILOQDSALQKKNCFQ
MYERMKLNNELKQLOQFOQSEVHNFDKLEEMSNVLOMOROHOSDLKMLAAKE
SOLREPOEAMATRESILADEKEPYSPPAKLTPKE"CRHRENDQIMNVEOWAKEQK
IANEKLGNKLEQVQKTIKLTGSKHLHNVNHLQBNKKLNEISEKKLKAGNPRIC
VKAPPNKLKPEFSQKGLCCALGWRGIPQDLTPKTDHKYMLPHSSGSSYC"
380..508
/misc_feature
  /gene="STAP"
  /note="LEUCINE ZIPPER MOTIF"
1341..1426
/misc_feature
  /gene="STAP"
  /note="LEUCINE ZIPPER MOTIF"
2891..2956
/misc_feature
  /gene="STAP"
  /note="LEUCINE ZIPPER MOTIF"
3369..3374
/polyA_signal
  /gene="STAP"
3396
/polyA_site
  /gene="STAP"
ORIGIN
Alignment Scores:
Pred. No.: 0.00201 Length: 3429
Score: 166.00 Matches: 137
Percent Similarity: 37.94% Conservative: 110
Best Local Similarity: 21.04% Mismatches: 225
Query Match: 5.16% Indels: 179
DB: 10 Gaps: 35
SEQ3_K705_COPY_650_1270 (1-621) x AB029919 (1-3429)
QY 17 AspValLysHisValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePhe 36
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 GACATCAAGAACTCTCACGATGTC-----TGCAGAGCGCAGGGGAAGACCTTG 250
QY 37 ArgLeuTyrSerValCysThrValGluLysAlaLeuIleLeuAlaArgGlnAsnLys 56
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 CAAGAGAACCCAGCTCTGGTG-----GAGGAAGCGATGCTGAAGCCACCAACAAG 304
QY 57 ArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrp 76
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 AAGCAAGCCCAAGCTTAGTATTTCACGCAACACAGATG-----343
QY 77 GlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer---SerGluThrProAsp 95
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 344 -----GACTTCGGGTCAATAAGCAATACCTCTCAGGCACTCCAGCAACTGAAGAAG 397
QY 96 SerGlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerIle 115
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 AAGCTGCTGACCTCCAGCAGGAACTG-----GAGTTCCGTACC---436
QY 116 LeuSerSerLysGlyGlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLys 135
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 437 -----CAGGAGCTCCAGGCTTCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 478
QY 136 HisAlaGlnGlyThrTyrSerSerAspSerThrLeu-----PheGly 149
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 TCGATCCTGGAGAAGCAGACTCTCCGACCTTCTTGTTCACCGTCACGTCAACAACCTTAAG 538

```

```

150 GluAspHisIleLysLeuSerAspGluGluSerProAsn-IlePheTrpSerLysLeuLe 169
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 GAAGATGAGGTCAATCTCTACGAGGAGGATGGGAAATCATAGTAAGACCGG----- 593
QY 169 uGlyGlyLysAsnProMetTrpLysTyrProSerAspThrProGlnArgAsnArgLysAr 189
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 594 ----GGGAGNAATCCATCTGGCGC---AAGACGACTTCCTTGGCGG-----635
QY 189 gValGlnTyrPheGluGlySerGluAlaSer-ProLysThrGlyAspGlyAsnAlaL 209
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 636 ----GGGACAAGATCCTCTCCCTAGA-----A 658
QY 209 yLysArgLysLysAlaSerAspValThrAsp-----ProArgValThrA 225
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 AGGAGCCTAAACCTCTACAGGGCAAGTACCACTTCCCTGAGCAACATAGAGTTACTG 718
QY 225 spProValAspAspAsp-GluArgLysAlaSer-----GlyLysAspHisMet 241
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 719 GAGTGCCAAGTGAAGATGCTGGAGGGGAACTCAGCGGACTCATTTGTCAGGACCTGAG 778
QY 242 GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThr 261
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 779 AACAAAGGAGACACCCGGAAGGTTTCAATCTACACTTCTCCGTGCGTGAATTCAAGAGCAT 838
QY 262 AspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIleSer 281
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 839 CAGGAGACCCCTA-----850
QY 282 GlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSerGln 301
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 851 -----AAGCGACTGCTGAAAGTCTGGCAAAAGGTTTCTGAACAGGATGAC 895
QY 302 ArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGln-----Val 318
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 896 -----CTAATCCAGGAGCTTCGGAATAAAGTGGCC-----TGCACCAACTCTTTGGTGC 943
QY 319 LeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 944 CTGGAGCGTGAGGAGCGGTTGATAAAATTACAAGCTGAGTTGCTTCTCTACACA---GCC 1000
QY 339 AsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAlaLeuSer 358
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1001 ACCCACCG-----CATCCCTCTCTCAGAAAGACTGTGAAGACATCAACAAG 1051
QY 359 TrpIleAlaAlaLeu-----363
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1052 ATCTGAAGCACTCTGCAAGACAGAAAGACAGCCAGTGCCTGCGTGGAGGAGTACCAG 1111
QY 364 ---LeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeu 382
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1112 AACCTTGTGAAGACCTGCGCATGGAGCTGGAAGCTGTGTGACAGACAGAAAGAAAGATC 1171
QY 383 AlaPheLysCysSerArgValGluValAspTyr-----IleTyrSerIleLeuSerCys 400
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1172 ATGAAGGACATGATGAAGCTGGAGCTAGATCTATATGCGGTGCGGGAAGAGACGTCTCTC 1231
QY 401 Met-----LysSerLeuPheLeuGluHis---ThrGlnGlyLeuGln 413
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1232 GTCATCGAAGAGACAGACAGGAGACTGTCTTCTGAGTACCGGCTGCAAGATCTGCAG 1291
QY 414 PheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGlu 433
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1292 ---CAACAGTACACAGAGTACACAGAGTGTCTCCTGAAGAAGGACAAGTTGCTCCAAGAC 1348
QY 434 SerLeuSerGlyAlaThrValArgAspGluLysIle-----AsnThrLys 448
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1349 -----AAGGATGAGAGCTGCACGAGCTGCACGAGTGAAGAATCTGATG 1387
QY 449 SerMetArgAsnSerSerGluAspGluGluCysMetThrGluLys---ArgCys-----465
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1388 CAGGTGCAGAACAGCCTCTCAGAGAAAGAGGAGAGCTGGAGAGCTACAGTGTACNACC 1447
QY 466 -----SerHisTyrSerThrAlaThrArgAspIleGlu 476

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n.model

Run on: July 11, 2005, 09:34:25 ; Search time 865 Seconds
(without alignments)

4249.895 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270

Perfect score: 3216
Sequence: 1 RADAFILFGSLNPSHDVKH.....SGNNKHFGSSNITQNAPDV 621

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5	
Zgapop 6.0 , Zgapext 7.0	
Deiop 6.0 , Delext 7.0	

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/ROBINSON929/runat_11072005_102151_29159/app_query.fasta_1.2325
-DB=N Geneseq 16Dec04 -QPMF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=ROBINSON929 @CGN 1 1 1234 @runat_11072005_102151_29159 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:*

1:	Geneseqn1980s.*
2:	Geneseqn1990s.*
3:	Geneseqn2000s.*
4:	Geneseqn2001as.*
5:	Geneseqn2001bs.*
6:	Geneseqn2002as.*
7:	Geneseqn2002bs.*
8:	Geneseqn2003as.*
9:	Geneseqn2003bs.*
10:	Geneseqn2003cs.*
11:	Geneseqn2003ds.*
12:	Geneseqn2004as.*
13:	Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3210	99.8	6571	4 AAA89354	Aaa89354 Arabidops
2	2977	92.6	10329	4 AAA89385	Aaa89385 Arabidops
3	2974	92.5	10329	4 AAA89353	Aaa89353 Arabidops
4	494.5	15.4	467	4 AAA89379	Aaa89379 Brassica
5	472	14.7	515	4 AAA89381	Aaa89381 Brassica

C	6	442	13.7	519	4	AAA89377	Aaa89377 Brassica
	7	166	5.2	5220	12	ADQ59166	Adq59166 MSI-H car
	8	166	5.2	5837	12	ADL12346	Adl12346 Human ste
	9	166	5.2	7764	5	ADL63336	Adl63336 Human ova
	10	166	5.2	7764	10	AD54628	Aad54628 Human chr
	11	166	5.2	7764	12	ADK60458	Adk60458 Angiogene
	12	166	5.2	7764	12	ADK60759	Adk60759 Angiogene
	13	166	5.2	7764	12	ADP73081	Adp73081 Angiogene
	14	164	5.1	3045	13	AD848467	Ad848467 Bacterial
	15	163	5.1	5858	2	AAQ10378	Aaq10378 Encodes h
	16	162.5	5.1	4769	12	ADQ64522	Adq64522 Novel hum
	17	161	5.0	5175	2	AAAX40200	Aax40200 Restin en
	18	161	5.0	5896	4	AAHS7404	Aahs7404 Human ske
	19	159	4.9	5857	6	ABL61968	Ab161968 Colon ade
	20	159	4.9	5857	10	ADB31363	Adb31363 Bicalutam
	21	159	4.9	5857	13	ADP56352	Adp56352 Human PRO
	22	159	4.9	6028	4	AAH72966	Aah72966 Human cer
	23	158	4.9	5943	4	AA822920	AA822920 DNA encod
	24	157.5	4.9	5883	12	ADQ87119	Adq87119 Human tum
	25	157.5	4.9	5883	13	ADQ84854	Adq84854 Human tum
	26	157.5	4.9	5883	13	ACN40924	Acn40924 Tumour-as
	27	157.5	4.9	6432	5	AA65903	AA65903 DNA encod
	28	157.5	4.9	7274	13	ADN60282	Adn60282 Human non
	29	157.5	4.9	7396	6	ABT08505	Abt08505 Human nov
	30	157.5	4.9	7396	12	ADO09868	Ado09868 Human NOV
	31	157.5	4.9	7516	4	AAK51987	Aak51987 Human pol
	32	156.5	4.9	6354	10	ADJ56352	Adj56352 Chicken c
	33	156	4.9	6364	8	ACA04019	ACA04019 cDNA down
	34	154.5	4.8	6965	10	ADD29780	Add29780 Human tum
	35	154.5	4.8	7636	10	ACF58150	Acf58150 Human gol
	36	154.5	4.8	7694	6	ABX04175	Abx04175 Human mRN
	37	154.5	4.8	7694	8	ABX10347	Abx10347 DNA encod
	38	154.5	4.8	7695	6	ABL68595	Ab168595 Kidney ca
	39	154.5	4.8	7695	10	ACF58151	Acf58151 Human gol
	40	154.5	4.8	7743	10	ACF58152	Acf58152 Human gol
	41	154.5	4.8	7761	10	ACF58153	Acf58153 Human gol
	42	154.5	4.8	8063	5	ADL63564	Adl63564 Human ova
	43	153	4.8	7985	4	ABL08391	Ab108391 Drosophil
	44	153	4.8	12120	4	ABL08390	Ab108390 Drosophil
	45	150	4.7	5847	6	AB232324	Abz32324 Candida a

ALIGNMENTS

RESULT 1
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
XX
AC
AC
AAA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
CDS 310..6315
FT /*tag= a
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-BP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
PA (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX

PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;

XX WPI: 2001-137952/14.

DR P-PSDB; AAB20062.

XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.

XX Claim 6; Page 22-32; 48pp; English.

XX The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg1). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-3' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAB89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression

XX Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8,13e-297 Length: 6571
Score: 3210.00 Matches: 620
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 99.81% Indels: 0
DB: 4 Gaps: 0

SEQ3_K705_COPY_650_1270 (1-621) x AAA89354 (1-6571)

QY	1	ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis	20
DB	2257	CGAGCTGATCGGTTATTTCTTTTGAAGCAGCTTGAATCCATCCGATGATGTTAAGCAC	2316
QY	21	ValGluLysIleLysIleGluSerCysSerGluArgThrIlyAsIlePheArgLeuTyrSer	40
DB	2317	GTAGAGAGATAAAAATCAGTGTCTCTGAAAGAACTAAGATATTCGATTTGTACTCA	2376
QY	41	ValCysThrValGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys	60
DB	2377	GTATGTACATTTGAAGAAAGCCCTGATTTCTGGCTAGGCAAAATATCGCGCAAAATAA	2436
QY	61	AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr	80
DB	2437	GCTGTAGAGAACCTAAACCGCTCTCTCAGCGACGCACTGCTCATGTGGGGGGCGTCATAC	2496
QY	81	LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe	100
DB	2497	TTATTTGATAAATCGATCATTTTTCACAGCAGTGAATCCAGATTCAGAGTTTCATTT	2556
QY	101	GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly	120
DB	2557	GAACATCTTATATGACGGCGGATTCATGAATTTCTCGTCCATACTTTCTTCCAAAGGT	2616
QY	121	GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr	140
DB	2617	GGAGAGAAAAATGAAGTGTGTCTACTTTTGGAGGCCAAGCATGCTCAGGGAAC	2676
QY	141	TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer	160
DB	2677	TACAGCAGTGTACTTCTATTTTGGTGAAGACCATATTAAATTTGTCAGATGAAGAGT	2736

QY	161	ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer	180
DB	2737	CCAAATATATTTTGGTCAAGAGCTTTGGGGGAAAAATCCTATGTGGAATACCTTCA	2796
QY	181	AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro	200
DB	2797	GATATCTCCCAAGGAATCGAAACGAGTTCAGTATTTTGGAGGGTCTCGAAGCGAGTCCC	2856
QY	201	LysThrGlyAspGlyGlyValAsnAlaLysLysAsgLysLysAlaSerAspValThrAsp	220
DB	2857	AAAACTGGCGATGGTGGAAATCCAAAGAACGAAAGAGGCTTCTGATGATGATCACTGAT	2916
QY	221	ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis	240
DB	2917	CCCCGGTCACTGATCCGCCAGTAGATGATGAAAGAAAGGCTCTCGGAAGGATCAC	2976
QY	241	MetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGly	260
DB	2977	ATGGGGGCTTTGGAGTCCAAAGATCAACACATCCAGTCAATGATGATAAATCTTCTGGT	3036
QY	261	ThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHisIle	280
DB	3037	ACAGATGGTACATTTGGATGGAAATGATGCTTTTGGCTTGATTTCTATGGGACCATATC	3096
QY	281	SerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIleProAspGluSer	300
DB	3097	TCTGAAATCCCAAGAGATATGTTAGTCAAGATTGGGGGAAAAATACCGGATGAATCA	3156
QY	301	GlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHis	320
DB	3157	CAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGGCAAACTTTGCCAAGTTTTCAT	3216
QY	321	LeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHis	340
DB	3217	CTTTTCAGATGCTTGACACAGCATGTCGGAAATTTTCTCGATATGTTATGAAATCAC	3276
QY	341	ArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIle	360
DB	3277	CGAATCTCAGAAAGAGCCAGCACTACTTTTCCAGGCATTCAGATAGCCCTGAGTTGGATT	3336
QY	361	AlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSer	380
DB	3337	GCAGCCTTGTGGTAAAGCAAAATCTTAGCCCAAAAGAAATCTCTGGTCGTCGCAATTC	3396
QY	381	GluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCys	400
DB	3397	GAATTAGCTTCAATGCTCTAGAGTAGAGTGGATATATTTATTCGATATTTGCTCTGC	3456
QY	401	MetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsn	420
DB	3457	ATGAAGAGTCTGTTCTCGAGCATACAAAGGTTTTCAGTTCGATTTGCTTGGTACTAAT	3516
QY	421	SerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThrVal	440
DB	3517	TCTAAACAGTCACTGGTGTAGCAAAACTAGTAATGAAAGTCTCTCAGGGGCTACAGTG	3576
QY	441	ArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGlyCysMet	460
DB	3577	CGTACGAAAGATTAATACGAAGTCGATCGCAATAGCTCAGAGGATGAGAGTGCATG	3636
QY	461	ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer	480
DB	3637	ACTGAGAAGAGATGTAGCCATTTATAGCACAGCAACAGAGATATCGAAAGACTATTAGT	3696
QY	481	GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys	500
DB	3697	GCATAAAAAAGAAATACAAAGAACAGTGCAAAAGCTTTGTACAGAGCATGGGAAAG	3756
QY	501	LysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSer	520
DB	3757	AAAAATGGAGCTGTTAAATATGATATGCAGACAAAGACGAAACTTGAAACTAGTAAAGCT	3816
QY	521	ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu	540

```

Db      3817 GTGAAGCAGCAGTAATTCGTTTACCTGTTTCACGGACCAGTACTCAAGTGGGTGATCTC 3876
Qy      541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
Db      3877 AAAGCTGCTGATCATATAATTATGAAGAAAGTTTGATGAATCAAAAGCTGAGAAAAATGAA 3936
Qy      561 CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGlu 580
Db      3937 TGCCCTCAAAAGTCTGGAGCAAAATGCACGAGTTTGCAAAGAAAGAGTTGGCTGAGGATGAA 3996
Qy      581 AlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCysValProIle 600
Db      3997 GCCTGTTGGATTATTCGATAAAGAGCTGGCGAGCTAAATATAAAGTTGTGTCCCAAT 4056
Qy      601 GlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIleSerGlnAsnAlaProAsp 620
Db      4057 CAAAGTGCACATAACAGCACTTTAGTGGTTTCATCAAAACATTTCCCAAAATGCTCCTGAT 4116
Qy      621 Val 621
Db      4117 GTA 4119

RESULT 2
AA889385
ID      AAA89385 standard; DNA; 10329 BP.
XX
AC      AAA89385;
XX
DT      04-NOV-2004 (revised)
DT      23-APR-2001 (first entry)
XX
DE      Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW      Gene silencing; silencing gene; MOM; ds.
XX
OS      Arabidopsis thaliana.
XX
FH      Key
FT      exon      1. .1008
FT      /tag= a
FT      /number= 1
FT      /tag= b
FT      /number= 1
FT      /tag= c
FT      /number= 2
FT      /tag= d
FT      /product= "Columbia protein"
FT      /note= "This sequence contains 15 introns."
FT      intron     2551. .2673
FT      /tag= e
FT      /number= 2
FT      /tag= f
FT      /number= 3
FT      /tag= g
FT      /number= 3
FT      /tag= h
FT      /number= 4
FT      /tag= i
FT      /number= 4
FT      /tag= j
FT      /number= 5
FT      /tag= k
FT      /number= 5.

FT      exon      3974. .4895
FT      /tag= 1
FT      /number= 6
FT      /tag= m
FT      /note= "ecotype Zurich"
FT      intron     4338
FT      /tag= n
FT      /number= 6
FT      /tag= o
FT      /number= 7
FT      /tag= p
FT      /number= 7
FT      /tag= q
FT      /number= 8
FT      /tag= r
FT      /number= 8
FT      /tag= s
FT      /number= 9
FT      /tag= t
FT      /note= "ecotype Zurich"
FT      intron     7481. .7615
FT      /tag= u
FT      /number= 9
FT      /tag= v
FT      /number= 10
FT      /tag= w
FT      /number= 10
FT      /tag= x
FT      /number= 11
FT      /tag= y
FT      /number= 11
FT      /tag= z
FT      /number= 12
FT      /tag= aa
FT      /number= 12
FT      /tag= ab
FT      /number= 13
FT      /tag= ac
FT      /number= 13
FT      /tag= ad
FT      /number= 14
FT      /tag= ae
FT      /number= 14
FT      /tag= af
FT      /number= 15
FT      /tag= ag
FT      /number= 15
FT      /tag= ah
FT      /number= 16

FT      exon      9001. .9211
FT      /tag= ag
FT      /number= 15
FT      /tag= ah
FT      /number= 16

XX      WO200100801-A2.
XX      04-JAN-2001.
XX      PD

```

XX 21-JUN-2000; 2000WO-BF005761.
 XX 23-JUN-1999; 99GB-00014623.
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX
 PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 XX
 XX WPI: 2001-137952/14.
 DR P-PSDB; AAB20063.
 DR
 XX Novel gene encoding a protein that controls gene silencing, in particular
 PT silencing of plant genes.
 PT
 XX Claim 5; Page; 48pp; English.
 PS
 CC The present sequence is that of an Arabidopsis thaliana ecotype Columbia
 CC gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino
 CC acid protein with similarity with ATPase/helicase proteins of the
 CC SWI2/SNF2 family, and which controls gene silencing, particularly plant
 CC gene silencing. Trans-acting modified loci were identified by T-DNA
 CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
 CC carries a heritable inactivated, methylated hygromycin resistance gene
 CC (hyg). Mutation of a silencing modified locus results in release of
 CC silencing of the hyg gene and restores hygromycin resistance. Plants
 CC homozygous for the silent resistance gene were subjected to
 CC transformation with a bar selectable marker gene under control of the T-
 CC DNA 1'-2' dual promoter. Transformsants were selected and their progeny
 CC screened for hygromycin resistance. The mutant phenotype (hygromycin
 CC resistance) was screened for genetic co-segregation with a specific T-DNA
 CC insert. Cloning of the tagged gene allowed characterization of the
 CC present sequence. 2 Allelic differences were found in the corresponding
 CC gene of A. thaliana ecotype Zurich, occurring at nucleotide positions
 CC 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful
 CC as a molecular tool for regulating gene expression. Note: the present
 CC sequence is not shown in the specification but is derived from the A.
 CC thaliana mutant silencer gene sequence given in AAA89353
 CC
 CC Revised record issued on 04-NOV-2004 : Correction to feature table key
 XX
 SQ Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3 59e-274 Length: 10329
 Score: 2977.00 Matches: 617
 Percent Similarity: 68.51% Conservative: 1
 Best Local Similarity: 68.40% Mismatches: 3
 Query Match: 92.57% Indels: 284
 DB: Gaps: 3
 SEQ3_K705_COPY_650_1270 (1-621) x AAA89385 (1-10329)
 QY 1 ArgAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
 DB 4172 CGAGCTGATGCGTTTATCTTTTGAAGCAGCTTGAATCCATCGCATGATGTTAAGCAC 4231
 QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrIlePheArgLeuTyrSer 40
 DB 4232 GTAGAGAGATAAAAATCGAGTCATGTTCTGAAGAAGCACTAAGATATTTCCGATTGTACTCA 4291
 QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys 60
 DB 4292 GTATGTACAGTTGAAGAAAAGCCCTGATTCTGGCTAGGCAAAATGAAGCGCAAAATAAA 4351
 QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 DB 4352 GCTGTAGAGAACCTAAACCGCTCTCTCACGACGCACTGCTCATGTGGGGGGCGTCATAC 4411
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 4412 TTATTTGATAAACTGGATCATTTTTCACAGCAGTGAAATCCAGATTCCAGGATTTTCATTT 4471

QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerIysGly 120
 DB 4472 GAACCAATCTATTATGGACGGCGTGATTCATGAATTCCTCGTCATACTTTCTTCCAAAGGT 4531
 QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 DB 4532 GGAGAAAGAAAATGAAGTCAAGCTGTCTACTTTTGGAGGCCAAGCATCTCAGGGAAT 4591
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 DB 4592 TACAGCAGTATTCTACTTCTATTGGTGAAGACCATATTAAAGTGTTCAGATCAGAGACT 4651
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSer 180
 DB 4652 CCAATATATATTTTGGTCAAAAGCTGTGGGGGAAAAAATCCCTATGTGGAATACCCCTTCA 4711
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerPro 200
 DB 4712 GATATCTCCCAAGGAATCGAAAAAGAGTTCAGTATTTTGAAGGGTCTCGAAGCGAGTCCC 4771
 QY 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp 220
 DB 4772 AAACTGGCGATGGTGGAAATCGAAGAACGAAAGAGGCTTCTGATGATCTCACTGAT 4831
 QY 221 ProArgValThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHis 240
 DB 4832 CCCCAGGTCACGTATCCGCGCAGTAGATGATGATGAAGAAAGGCTCTGGGAAGGATCAC 4891
 QY 241 Met----- 241
 DB 4892 ATGGGTAAATAATAGTTAATTTCTGTCGATACCTCTAGTGTTCATTTGATATGCAACTA 4951
 QY 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
 DB 4952 CTTTGTGATATCTTTCTCAGAGGGCT-TTGGAGTCACCAAAAGTCATACACTCCAG 5010
 QY 254 SerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu 273
 DB 5011 TCATCATGTAAATCTTCTGGTACAGATGATGATGATGATGATGATGATGATGATGATG 5070
 QY 274 TyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp 293
 DB 5071 TATTTCTAGGGCAGGCATATCTCTGAATCCAGAGGATATGTTAGTCTAGTCAAGATTGG 5130
 QY 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
 DB 5131 GGGAAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTAAAGCCGAAGATGGCA 5190
 QY 314 LysLeuCysGlnValLeuHisLeuSer----- 322
 DB 5191 AAACITTTGCCAAGTTTTCATCTTTC-AGTAAAGTGGCCTTTTTCACCTCCACAACATTAT 5249
 QY 322 ----- 322
 DB 5250 TTAGCCTTCATATGCTTATATATATAGCTGATGTCACCTGATGTTTACCTGATTTCCCTG 5309
 QY 322 ----- 322
 DB 5310 TTACAGCCAAATGTGAGAGTTTTTATCTTCAACTATATCCATCCGTTTAAGCATATTTTA 5369
 QY 322 ----- 322
 DB 5370 TTTCTTATATCTGGCTTCGTTACCAATGCACTGTTAAATGAGCAACTGCTGCACAAAAC 5429
 QY 322 ----- 322
 DB 5430 AGTAGGTAGTTATGTGCCTCATGTCATTCATTGTTTATTGAAGCAAGAAAATTTCTGTCT 5489
 QY 322 ----- 322
 DB 5490 ACTTTACATGATCCATCTCTGGGAGTATATATACTATATATAACCTTTAGGCCCTTTGTACCT 5549

FT		/tag= m
FT	note= "ecotype Columbia"	
FT	intron	4896..4975
FT		/tag= n
FT		/number= 6
FT	exon	4976..5217
FT		/tag= o
FT		/number= 7
FT	intron	5218..5777
FT		/tag= p
FT		/number= 7
FT	exon	5778..5882
FT		/tag= q
FT		/number= 8
FT	intron	5883..6082
FT		/tag= r
FT		/number= 8
FT	exon	6083..7480
FT		/tag= s
FT		/number= 9
FT	allele	6721
FT		/tag= t
FT	note= "ecotype Columbia"	
FT	intron	7481..7615
FT		/tag= u
FT		/number= 9
FT	exon	7616..7771
FT		/tag= v
FT		/number= 10
FT	intron	7772..7914
FT		/tag= w
FT		/number= 10
FT	exon	7915..8070
FT		/tag= x
FT		/number= 11
FT	intron	8071..8153
FT		/tag= y
FT		/number= 11
FT	exon	8154..8318
FT		/tag= z
FT		/number= 12
FT	intron	8319..8451
FT		/tag= aa
FT		/number= 12
FT	exon	8452..8629
FT		/tag= ab
FT		/number= 13
FT	intron	8630..8718
FT		/tag= ac
FT		/number= 13
FT	exon	8719..8918
FT		/tag= ad
FT		/number= 14
FT	intron	8919..9000
FT		/tag= ae
FT		/number= 14
FT	exon	9001..9211
FT		/tag= af
FT		/number= 15
FT	intron	9212..9284
FT		/tag= ag
FT		/number= 15
FT	exon	9285..10329
FT		/tag= ah
FT		/number= 16
XX	WO200100801-A2.	
PN		
XX	04-JAN-2001.	
XX		
Pf	21-JUN-2000; 200WO-BF005761.	
XX		
PR	23-JUN-1999; 99GB-00014623.	

Db 4532 GGAGAGAAATGAAGTCAAGCTGTCTACTTTTGGAGGCGAAGCATGCTCAGGAACT 4591
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 4592 TACAGCAGTGAATCTACTCTATTGGTGAAGACCATATTAAAGTTGTTCAGATGAAGAGT 4651
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpProSer 180
Db 4652 CCAAAATATATTTGGTCAAAAGCTGTGTGGGGGAAAAAATCTATGTGGAATAACCCTTCA 4711
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTrpPheGluGlySerGluAlaSerPro 200
Db 4712 GATACTCCCAAGGAATCGAAACGAGTTTCAGTATTTTGGGTTCTGAAGCGAGTCCC 4771
Qy 201 LysThrGlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAsp 220
Db 4772 AAAACTGGCGATGTGTGAATGCAAGCAAGCAAGCAAGGCTTCTGATGATGCTACTGAT 4831
Qy 221 ProArgValThrAspProValAspAspGluArgLysAlaSerGlyLysAspHis 240
Db 4832 CCCGGGTCACTGATCCGCCAGTAGATGATGATCAAAAGAAAGGCTCTGGGAAGGATCAC 4891
Qy 241 Met----- 241
Db 4892 ATGGTAAATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTTGATATGCAACTA 4951
Qy 242 -----GlyAlaLeuGluSerProLysValIleThrLeuGln 253
Db 4952 CTTTGTGCTACTATCTTCTACAGGGGCT-TTGGAGTCAACAAAGTCATAACCTCCAG 5010
Qy 254 SerSerCysLysSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeu 273
Db 5011 TCATCATGTAAATCTTCTGTGTACAGATGTGTACATTTGGATGGAATGATGCTTTTGGCTTG 5070
Qy 274 TyrSerMetClySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrp 293
Db 5071 TATTTATGGGAGCCATATCTCTCGAATCCAGAGGATATGTAGTACTCAAGATTGG 5130
Qy 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAla 313
Db 5131 GGGAAATACCGATGATCAACAGAGAGGCTCCACACTGTTTAAAGCCGAGATGGCA 5190
Qy 314 LysLeuCysGlnValLeuHisLeuSer----- 322
Db 5191 AAACTTTGCCAAGTTTGTGATCTTTC-AGTAAAGTGGCCTTTTTCACCTCCCAACTTATT 5249
Qy 322 ----- 322
Db 5250 TTAGCCTTGCATATGCTTATATATAGCTGATTCGAACTGTAGTTGTATCCTGATTTCTTG 5309
Qy 322 ----- 322
Db 5310 TTACAGCAATGTGAGAGTTTATCTTCAACTATATCATCCGTTTAAGCATATTTTA 5369
Qy 322 ----- 322
Db 5370 TTTCTTATATCTGGCTTCGTTACCAATGCACCTGTAAATAGCAACTGCTGCACAAAC 5429
Qy 322 ----- 322
Db 5430 AGTAGGTAGTTATGTGCTCATGCTATTCATTTGTTTATGAAGCAAGAAATTTCTGTCT 5489
Qy 322 ----- 322
Db 5490 ACTTTACATGATCCATCTGTGGAGTATATAACTATATATAACCTTAGGCCCTTTGTACCT 5549
Qy 322 ----- 322
Db 5550 GGCTGATCAAGACATGTCAAAAGTTTATCTGTTCGCTGTGTGATAGAAACTAATACAG 5609
Qy 322 ----- 322

Db 5610 TGCTGTATGCTATTTTAAAGGTAGTCTTATGTCTTTCACATATTGGCTAATAGATGTTTCCG 5669
Qy 322 ----- 322
Db 5670 CTGTGCTGCCATATACTTCTGTGATTATCACGGTGTCTCGTCTATCAAAATTTGTACTAA 5729
Qy 323 -----AspAlaCysThr 326
Db 5730 AAGGTATTTTGCATATGTGTGATTGGTTAAACAGATTATTTTGTTCAGGATGCTTGCACA 5789
Qy 327 SerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluPro 346
Db 5790 AGCATGGTGGAAATTTTCTCGAATATGTATTGAAATCACCAGATCTAGAAAGACCA 5849
Qy 347 AlaThrThrPheGlnAlaPheGlnIleAlaLeu----- 357
Db 5850 GCCACTACTTTTTCAGGCATTTCCAGATAGCCCT-GGTATGACAGCATTTTACTTTGATAATT 5908
Qy 357 ----- 357
Db 5909 TATGCATTTGTTTCTTTCATCATCTGCTTTTGTAGAAATGTCTCTCAGAAGGCAGACTCC 5968
Qy 357 ----- 357
Db 5969 TTTAGTTTAACTTTTCCAATCATAGGATTCAAATATCCATTAAGTGGCCTTTTGATCGCTG 6028
Qy 358 -----SerTrp 359
Db 6029 CATAATATATGAATAGTTGACATACCTGAATACCTTGTTAATAATGCAATTTTCAGAGTTGG 6088
Qy 360 IlAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsn 379
Db 6089 ATTGCAGCCTTGTGTGAAGCAAAATCTTAGCCACAAAGAAATCTCTGGTCCGTGCAAA 6148
Qy 380 SerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSer 399
Db 6149 TCTGAATTAGCTTTCAATGTCTCTAGATAGAGGTGATATATTTATTCGATATTGTCC 6208
Qy 400 CysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 6209 TGCATCAGAGTCTGTTCTGTGAGCATACACAGGTTTGCAGTTTCGATTTGTTGTACT 6268
Qy 420 AsnSerLysGlnSerValSerThrLysLeuValAsnGluSerLeuSerGlyAlaThr 439
Db 6269 AATTTCTAAACAGTCAGTGTGTAGCACAATACTAGTAAATGAAAGTCTCTCAGGGGTACA 6328
Qy 440 ValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAspGluGluCys 459
Db 6329 GTCGCTGACGAAAGAAATTAATACGAAGTCGATCGAAATAGCTCAGAGGATGAAGAGTGC 6388
Qy 460 MetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIle 479
Db 6389 ATGACTGAGAGAGATGTAGCCATTATAGCAGCAACCAAGAGATATCGAAAGACTATT 6448
Qy 480 SerGlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGlu 499
Db 6449 AGTGGCATAAAAAGAAATACAAAGCAAGTGCATAAAGCTTGTACAAAGAGCATAGGAA 6508
Qy 500 LysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLys 519
Db 6509 AAAAAATGGAGCTGTTAAATATGTATGCAGCAACAAAGCAAGAACTTGAACCTAGTAAA 6568
Qy 520 SerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp 539
Db 6569 AGTGTGGAGCAGCAGTAAATTCGTATTACTGTTCACGACCAAGTACTCAAGTGGGTGAT 6628
Qy 540 LeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsn 559
Db 6629 CTCAAACTGCTGATCATAAATATGAAGAAAGTTTGTATGAATCAAAAGTGAAGAAAT 6688
Qy 560 GluCysLeuLysSerLeuGlnMetHisGluValAlaLysLysLysLeuAlaGluAsp 579
Db 6689 GAATGCTTCAAAAGTCTGGAGCAAAATGCACGAGGTTTGCAAAGAAAGAAAGTTGGCTGAGGAT 6748

Alignment Scores:

Pred. No.: 1.03e-35 Length: 515
 Score: 472.00 Matches: 100
 Percent Similarity: 60.87% Conservative: 26
 Best Local Similarity: 48.31% Mismatches: 39
 Query Match: 14.68% Indels: 42
 DB: 4 Gaps: 2

SEQ3_K705_COPY_650_1270 (1-621) x AAA89381 (1-515)

QY 387 SerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeu 406
 Db 1 TCTAGCAAGAGGTGGATTATGATATCTTTCTGCTACTGCATGAAGATCTATTCGTG 60
 QY 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
 Db 61 GGGCGCACACAGGTTTCCAA----- 81
 QY 427 SerThrLysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn 446
 Db 81 ----- 81
 QY 447 ThrLysSerMetArgAsnSerSerGluAspGluCysMetThrGluLysArgCysSer 466
 Db 82 -----GAAAGGGTGAAGATGCTGCTGAGAAAGAGGTAGC 120
 QY 467 HisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysTyr 486
 Db 121 CATTATAGCTCAGTACCAGAGGTGTGAAGACGATATTAGCGACATCAAAAGAAATGC 180
 QY 487 LysLysGlnValGlnLysLeuValGlnGluHisGluLysLysMetGluLeuLeuAsn 506
 Db 181 AGTAAGAGCGCTGCATAGCTTGTACAAACCTCGAGGAGAGAAAGATGGACCTGATGAAT 240
 QY 507 MetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaIleValIle 526
 Db 241 AGGAATGCTCTCAAGAGCAGGAACTTCAGAAATGTGAAGAGGTGAAGCATCATTTATT 300
 QY 527 ArgIleThrCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAsp 544
 Db 301 CGTGTACCTATTTCAGGTATAATATCTCAGAGGTACATGATGCTCTCCAAACGCTGGAA 360
 QY 545 HisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSer 564
 Db 361 TGTACTTTTGAAGAAAGTTTGATGATCTCAAAGGAGAGTTGATGAATGCCTTGAAGT 420
 QY 565 LeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrpIle 584
 Db 421 TTAGAGCAATAAACCGAGGCTGGAAAGAGAAAGTTGGCTGAAGATGAAGCCTTGTGATT 480
 QY 585 AsnArgIleLysSerTrpAla 591
 Db 481 AGTCGGATAGAGAAATGGGCA 501

RESULT 6

AAA89377/c
 ID AAA89377 standard; DNA; 519 Bp.

AC AAA89377;

DT 11-SEP-2003 (revised)

DT 23-APR-2001 (first entry)

XX Brassica oleracea silencing gene homologous DNA seq1-23.

XX Gene silencing; silencing gene; MOM; ds.

XX Brassica oleracea; acephala.

XX WO200100801-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.
 XX 23-JUN-1999; 99GB-00014623.
 XX (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 XX WPI; 2001-137952/14.
 XX Novel gene encoding a protein that controls gene silencing, in particular
 XX silencing of plant genes.
 XX Example 6; Page 44; 48pp; English.
 XX This is the sequence of Seq1-23, a genomic DNA sequence of Brassica
 XX oleracea acephala that shows homology to the newly identified silencer
 XX gene (see AAA89353) of Arabidopsis thaliana. Seq1-23 is a partial
 XX sequence of clone 1, which was isolated from a genomic library screened
 XX with MOM cDNA. Partial sequences of clone 1 (see also AAA89378-80) showed
 XX similarity to different regions of the MOM gene (80-86% at DNA level and
 XX 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-
 XX terminal parts of the MOM protein (see AAB20062). Gene silencing is
 XX useful as a molecular tool for regulating gene expression. (Updated on 11
 XX -SEP-2003 to standardise OS field)
 XX SQ Sequence 519 Bp; 150 A; 117 C; 101 G; 149 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 7.91e-33 Length: 519
 Score: 442.00 Matches: 89
 Percent Similarity: 84.82% Conservative: 6
 Best Local Similarity: 79.46% Mismatches: 13
 Query Match: 13.74% Indels: 4
 DB: 4 Gaps: 1

SEQ3_K705_COPY_650_1270 (1-621) x AAA89377 (1-519)

QY 1 AtqAlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHis 20
 Db 324 CGCGCTGAGTCTTTTATCTTTTGGAGCAGCTGAATCCATCGCATGACGTTAAGCTC 265
 QY 21 ValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSer 40
 Db 264 TTAGAGAGATAAAGGTTGAGTCGTCTCTGAAGAACTAAAGATTTTCCGTTGTACTCA 205
 QY 41 ValCysThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLys 60
 Db 204 GTGTTTACAGTTGAAGAAAAAACCTGATTCTTGCT-----AGGCAAAATAAG 157
 QY 61 AlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 156 CCTCTAGAGAACCTTAACCGTCTCTCACATGTCATTTGCTGTTGGGGGGCATCATAT 97
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 96 TTATTTGATAAGCTGGACACATCTCCATGGGAGTGGAAACCCAGGTTCCAGGATTCATT 37
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPhe 112
 Db 36 GAACAGTCTATTAAAGGATCCGTAANNCCAGGAATTC 1

RESULT 7

ADQ59166

ID ADQ59166 standard; cDNA; 5220 BP.

XX ADQ59166;

XX 09-SEP-2004 (first entry)

XX MSI-H carcinoma cDNA sequence SEQ ID NO:3.

Db 2760 GCGGACACAGTCGTCTATCTTTGACTCTGACTGTAACCCCGAAGTACTTGCAGGCACAA 2819
Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
Db 2820 GCCCGAGCCCATAGATTGGTCAGAGAGACAGGTAAATATTACCGCTTAGTTACAAAG 2879
Qy 42 CysThrValGluGluLysAla 49
Db 2880 GCGAGCTGTGGAGGAGAGATCATAGAACGGGCGCAAAAGAGAGATGTTATTAGATCATCTG 2939
Qy 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn 64
Db 2940 GTGATTACAGCGATGACACCACTGGCCGGAGCATCTCGAAACAACTCAGGAAGGTCC 2999
Qy 65 3000 AACTCAATCCTTTTAATAAGAGAGAGCTGACAGCTATTGTAATTTGGAGCAGGAGAT 3059
Db 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Qy 3060 CTCTTCAAGAACTGGAA 3089
Db 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3090 GAACCTCAGGAATGGAT 3134
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3135 ACGAGAGAGAAAGAGTGTCAACAGTGCACAGATGAATCTTA 3191
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3192 GTTGCCAACTTTGCCAATCGAAGTGAAGTGAAGAGAGCTA 3239
Qy 161 ProAsnIlePheTyrSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 3240 CCTCACAAGACTGGGATGAGATCATT 3266
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
Db 3267 3320 CCAGAGNACAAGGAAAGAAAGTAGAGGAGAGAGCGGCGAGAGGAGGTAGAA 3320
Qy 195 3321 GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTTAAAGAGCTCAGACAAATGACAGT 3380
Db 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
Db 3381 GACTCTGACACTGAGTCTAAGAGGCGAGCCCGAGAGATCCTCTGCTTCTGAGAGTGAAACG 3440
Qy 224 ThrAspProProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 3441 GAAGAC 3464
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 3465 3515 CCAAGCGCAGAGGCGCTCCGAGAGTGTGCGGAAGGACCTCGTGGAGGGA 3515
Qy 264 ThrLeuAspGlyAsn 278
Db 3516 TTTACTGATCGAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAAGTTT 3563
Qy 279 HisIleSerGlyIleProGluAspMetLeu 296
Db 3564 3611 GGCTCCCTCTTGAACGGCTGGAGTCATAGCAGCTGAT 3611
Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 3612 GTAGATAAGTCGTGGCAGATCTG 3647
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 3648 GAACGTATCCAC 3701

Qy 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 3702 AAAGAAATATGCCAGCGAGGAAAGGACCCAGGAAAGG 3758
Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 3759 ATATCCGGAGTT 3812
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal 394
Db 3813 3860 GATGCTGCATTAATCTATCCCTGTGGACCCCTGAGCAAAAAA 3860
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 3861 TACTCTTAACCTGTCGTCAAAGCTGCACATTTTGTAGTAGTGGGGGTGGAAGAT 3920
Qy 415 AspCys 422
Db 3921 GATTCCTGCCTGTGCTGGGATTTTGAACATGGCTATGGAACCTGGGAGTTAATAA 3980
Qy 422 sGlnSerValSerThrLysLeuValAsn 432
Db 3981 ACAG 4028
Qy 433 4029 GATAAAAGCCTCAGGGGAGCAGCTACAGACCCGAGCGGATTACTTGTGAAGCTGCTC 4088
Qy 451 ArgAsnSerSerGlu 463
Db 4089 AGAAAGGCTGTGGAGAAGAAGGGGCTGTGCACAGTGGGGAGAGGCCCAATTAAGAAG 4148
Qy 464 Arg 464
Db 4149 CGGAAGCCTCGGTTAAAGAGGAAACAAAGTGCCCGCTGAAGAGGAGCATGGAAAT 4208
Qy 465 4209 GAGCTTTTCATCTCTAGGCAATTCAGATAATCCATCAGAAGAGGAGAGTGAAGATGAT 4268
Qy 481 GlyIleLysLysLysTyrLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4269 GCGTTGGAAAAAGTCCAATGAAAAAACAACAAG 4325
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys 511
Db 4326 AAGGAGAAACAATGAGTTCTAGGAAAGACAAGAAGGGGACAGGAAGAAGAAGTCA 4385
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla 523
Db 4386 AAAGATAAGAAAGAGAGCCTAAAAGTGGTGTGATGTCCTCAATCTTCGAGTAAATCAAAGCGA 4445
Qy 524 4446 TCTCAGGCTCTGCTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGAGATGAG 4505
Db 539 4506 GATGATGATCTGGACCAGGAGACATTCAGCATATGTAGGAGAGGATGAGGCCCTGAAA 4565
Qy 540 4566 AAGGCACCTGAAACAGCTCGACAAACCTGACAAAGGGGCTCAACGTCGAACACAGCTGGAA 4625
Db 553 4626 CACACCCGAACTGCTGCTGAAATCGAGAGACCGGATAGCCGAGTGCCTTAAGCCTAC 4685
Qy 564 564
Db 4686 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGGATTTTGTTCCAAGTTT 4745

QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
 Db 4746 ACAGAATTTGATGCTCGAAACTGCATAAGTTATACAGATGGCTCATAGAAAGGTCT 4805
 QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
 Db 4806 CAGAGAGAGAG-----GAGCAAAAGAGAAAGAGGAC 4838
 QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
 Db 4839 GTG-----ACTGGGGTAAAGAAACCATTTTCGTCAGAGGCTCAGGCTCCAGCGG 4889
 QY 614 -----IleSerGlnAsn 617
 Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 9

ADL63336

ID ADL63336 standard; DNA; 7764 BP.

XX

AC ADL63336;

XX

XX 20-MAY-2004 (first entry)

XX

DE Human ovarian cancer DNA marker #21548.

XX

KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.

XX

OS Homo sapiens.

XX

XX WO200170979-A2.

XX

XX 27-SBP-2001.

XX

XX 21-MAR-2001; 2001WO-US009126.

XX

XX 21-MAR-2000; 2000US-0191031P.

XX

XX 25-MAY-2000; 2000US-0207124P.

XX

XX 15-JUN-2000; 2000US-0211940P.

XX

XX 07-JUL-2000; 2000US-0216820P.

XX

XX 25-JUL-2000; 2000US-0220661P.

XX

XX 21-DEC-2000; 2000US-0257872P.

XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

XX Lee J, Lillie J;

XX

XX WPI; 2001-611502/70.

XX

XX Disclosure; SEQ ID NO 21548; 106pp; English.

XX

PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

PS The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein

CC fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention.

XX
 SQ Sequence 7764 BP; 2342 A; 1534 C; 1892 G; 1979 T; 0 U; 17 Other;

Alignment Scores:

Pred. No.: 0.000127 Length: 7764
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conservative: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x ADL63336 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
 Db 3342 GCGGACACAGTCGTCATCTTGACTCTGACTGGAAACCCAGAGTACTTGACGGACAA 3401
 QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuLysSerVal 41
 Db 3402 GCCGAGCGGCATAGAAATGGTCAGAAAGACGAGTAATATTTACCGCTTAGTACAAAG 3461
 QY 42 CysThrValGluGluLysAla-----Leu 49
 Db 3462 GGGAGTGTGGAGGAGGAGATCATAGAACGGCGCCAAAGAGAGTGGTATTAGATCATCT 3521
 QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
 Db 3522 GTGATTGAGCGCATGGACACACACTGGCCGAGCAGTCTCGGAAACAACTCAGGAAGTCC 3581
 QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 3582 AACTCAATCCTTTTAAATAAGAGAGAGCTCACAGCTATTTTGAAATTTGGAGCAGGAGT 3641
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 3642 CTCCTCAAGAGAACTGGAA-----GGGGAGGAGATCA 3671
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
 Db 3672 GAACCTCAGGAATGGAT-----ATAGATCAATTTTGGCGTTGGCTGAA 3716
 QY 121 GlyGluGluAsnGluValLysLeuCysLysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 3717 ACCAGAGAGAATGAAGTGTCAACAGTGAACAGATGAATCTCTA---TCACAGTTTAAG 3773
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
 Db 3774 GTTGCCAACTTTGCAACAAATGGAAGATGAAGAGAGCTA-----GAAGAGCGT 3821
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 Db 3822 CCTCACAAGGACTGGATGGATGATCAT-----3848
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 Db 3849 -----CCAGAGGAACAAAGAAAGTAAGTAGAGGAGGAGCGGAGAGGAGGTAGAA 3902
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3903 GAAATTTATATGCTGCTCGAATTCGGAGTTCACACTAAAGAGGCTCAGACAAATACAGT 3962

3522 GTGATTGAGCGGATGGACACCACTGGCCGGAGCAGATCTCTGGAACAACTCAGGAAGTCC 3581
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db AACTCAAACTCTTTTAATAAAGAGAGCTGACAGCTATTTTGAATTTGGAGCAGAGGAT 3641
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db CTCTTCAAGAACTGGAA-----GGGAGGAATCA 3671
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db GAACCTCAGCAATGAT-----ATAGATCAAAATTTTGGTGGCTGAA 3716
QY 121 GlyGluGlnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db ACGAGAGAGAAATGAAGTGTCAACAGTGCACAGATGAACTTCTA--TCACAGTTTAAG 3773
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db GTTGCCAACTTTGCAACAATGGAAGATGAAGAAGAGCTA-----GAAGAGCGT 3821
QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db CCTCACAAGACTGGGATGAGATCAT-----3848
QY 181 AspThrProGlnArgAsnArgLysArgValGlnThrPheGlu-----194
Db -----CCAGAGGAACAAGAAAGTAGAGGAGGAGGAGCGGCGCAGAGGAGTAGAA 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
Db GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTAAAGAGGCTCAGACAATGACAGT 3962
QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
Db GACTCTGACACTGAGTCTAGAGGAGGCGCCAGAGATCCTCTCTCTGAGAGTGAACG 4022
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db -----TCTGATGATGACAAGAAG-----4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db -----CCAAAGCGCAGAGCGCTCCAGAGAGTGTGCGGAAGCACTCTGCGAGGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db TTTTACTGATGAGAGATCCGAAGGTTTCATCAGGCTTATAAGAAGTTT-----4145
QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db -----GGTCTCCTCTTGAACGGCTGGAGTGTCTAGCAGTGTAT--GCTGAGCTG 4193
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db GTAGATAAGTCTGGTGGCAGATCTG-----AAGCGCTGGGT 4229
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db GAACTATTCAC-----AACAGCTGTGTTCAGCAATGAGGAATATGAAGAGCAGCTG 4283
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db AAGAANAATGCCAGCGGGAAGGACCAAGGGAAGG-----AGAGGTCCAACAATCAAG 4340
QY 357 LeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db ATATCCGGAGTT---CAGGTTAATGTGAATCATTCATTCACATCAAGAGGAGTT---4394
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db -----GAGATGCTGCATTAATCTATCCCTGTGACCGCTCAGAGAAAAA 4442
QY 4395 -----

395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db TACTGCTTAACCTGCTGTGTCAAAGCTGCACATTTTGTGTAGAGTGGGGGTGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
Db GATTTCTCGCTGTGCTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATTAAA 4562
QY 422 sGlnSerValValSerThrLysLeuValAsn-----432
Db -----ACCAGAGCTTAATTAATCTGACAAAAATTCGCCGGTGGAGACA 4610
QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
Db GATAAAAGCCTCAGGGGAAGCAGCTACAGACCGAGCGGATTAATTTGTAAGCTGCTC 4670
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db AGAAAGGGTCTGGAGAAGAGGGGGCTGTACAGGTGGGGAGGAGGCCAAATTAAGAAG 4730
QY 464 Arg-----464
Db CGGAAGCCTCGGTAAGAAGAAACAAGTCCCGAGCTGAAGAGGAGCATGGAATT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db GAGCTTTCATCTCTAGGCATTCAGATAATCCATCAGAAGAGGAGGAGTGAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluLys 500
Db GCGTTGGAAAAAGTCCAAATGAANAANAACAAGAG--AAGAAAGAGAACAAGGAGAAC 4907
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----511
Db AAGGAGAAACAATGAGTTCTAGGAAGAACAAAGAGGGGACAGAGAAAGAAAGAGTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----523
Db AAGATAAAGAAAGAGAGGCTTAAAGTGTGTGATGCCAAATCTTCGAGTAAATCAAAGCGA 5027
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----539
Db TCTCAGGGTCTGTCCATATTACAGCAGGAAGTGAACCTGTCCCAATTCGAGAGGATGAG 5087
QY 539 -----539
Db GATGATGATCTGACAGAGAGACATTCACCATATGTAAGAGAGGATGAGGCCCTGAAA 5147
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db AAGGCACCTGAACAGCTGCACAAACCTGACAAGGGGCTCAACGTGCAAGAACAGCTGGAA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer-----564
Db CACACCCGAAGTGGCTGTGTAATTCGAGAGACCGGATAGCCGAGTGCCTTAAAGCTTAC 5267
QY 564 -----564
Db TCAGATCAGAGACACATCAAACTCTGGAGGAGAACCTATGATTTTGTTCCTCAAGTTT 5327
QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db ACAGAAATTTGATGCTCGAAAACTGCATAAGTTATACAAGTGGCTCATAGAAGAGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
Db CAAGAAGAGAG-----GAGCAAAAGAGAGACGAC 5420
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db -----ACTGGGGGTAAAGAAACCATTTTCGTCCAGAGGCTTCAGCTCCAGCCGG 5471
QY 5421 GTG-----

QY 614 -----IleSerGlnAen 617
DB 5472 GACTCTCTGATATCTCAGTCC 5492

RESULT 12

ADK60759
ID ADK60759 standard; DNA; 7764 BP.

AC ADK60759;

DT 06-MAY-2004 (first entry)

XX DE Angiogenesis differentially expressed gene GS-N32.

XX ds; vasotropic; antiarthritic; hypotensive; antianigmal;
KW antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;
KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;
KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
KW ischemia; angina; myocardial infarction; chronic heart disease;
KW cardiac congestion; macular degeneration; osteoporosis.

XX OS Homo sapiens.

XX PN FR2836686-A1.

XX PD 05-SEP-2003.

XX PF 04-MAR-2002; 2002FR-00002717.

XX PR 04-MAR-2002; 2002FR-00002717.

XX PA (GENE-) GENE SIGNAL.
XX PA (ALMA/) AL MAHMOOD S.

XX PI Colin S, Schneider C, Al Mahmood S;

XX DR WPI: 2004-013911/02.

XX DR P-PSDB; ADK60509.

XX PT Compositions containing nucleic acid or polypeptide differentially
PT expressed in angiogenesis are useful to diagnose, prognosis and treat
PT angiogenic disorders including tumor vascularization and heart disease.

XX PS Claim 2; SEQ ID NO 33; 405pp; French.

XX CC The invention relates to a novel pharmaceutical composition active on
CC angiogenesis comprising an endothelial cell nucleic acid whose expression
CC is induced by an angiogenic factor and inhibited by an angiostatic agent
CC or its complement or fragment, a polypeptide sequence encoded by the
CC nucleic acid or its fragment, a molecule capable of inhibiting expression
CC of the nucleic acid or a molecule which binds to the polypeptide
CC sequence. The invention is used to diagnose, prognosis or treat an
CC angiogenic disorder in a mammal, particularly a human. The disorder is
CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
CC endometriosis associated with neovascularization, restenosis due to
CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
CC myocardial infarction, chronic heart disease, cardiac congestion or
CC macular degeneration due to age or osteoporosis. This sequence
CC corresponds to a differentially expressed DNA used in the composition of
CC the invention.

XX SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000127 Length: 7764

Score: 166.00 Matches: 162

Percent Similarity: 33.9% Conservative: 113

Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 12 Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x ADK60759 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAenProSerHisAspValHisVal 21
DB 3342 GCGGACACAGTCGTCTCTGACTCTGAGTCCGACCCAGATGACTTGCAGGCACAA 3401
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
DB 3402 GCCCGAGCGCATAGAAATTGGTCAGAAGAAGCAGGTAATAATATTACCGCTTAGTTACAAG 3461
QY 42 CysThrValGluGluLysAla-----Leu 49
DB 3462 GGGACTGTGGAGGAGAGATCATAGACGGGCCAAAGAGAGATGGTATTAGATCATCTG 3521
QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAen-----64
DB 3522 GTGATTACGCGCATGACACACCACTCGCGGACGATCCTCGMAAACAACACTCAGAAAGTCC 3581
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyr 80
DB 3582 AACTCAAAATCCTTTTAATAAAGAGAGCTGACAGCTATTTGAAATTTGGACGAGGAT 3641
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerSerGlyValSerPhe 100
DB 3642 CTCCTTCAAGAACTCGAA-----GGGAGGAATCA 3671
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
DB 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTGGCGTTGGCTGAA 3716
QY 121 GlyGluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
DB 3717 ACGAGAGAGATGAAGTGTCAACAAGTGCACAGATGAATTTCTA---TCACAGTTTAAG 3773
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer 160
DB 3774 GTTGCCAACTTTGCAACAATGGAAGTGAAGAGCTA-----GAAGAGCGT 3821
QY 161 ProAsnIlePheTyrSerLysLeuGlyLysAsnProMetTyrLysTyrProSer 180
DB 3822 CCTCACAAGGACTGGGATGAGTATC-----3848
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
DB 3849 -----CCAGAGGAACAAGGAAAGTAGAGGAGAGAGCGCAGAGGAGCTAGNA 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
DB 3903 GAAATTTATATGTCCTCGAATTCGAGTTCCATAAAGGCTCAGACAAATGACAGT 3962
QY 204 AspGlyGlyAsnAlaLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
DB 3963 GACTCTGACACTGAGTCTTAAGAGGAGCGCCAGAGATCCTCTGCTTCTGAGAGTGAACG 4022
QY 224 ThrAspProProValAspAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
DB 4023 GAAGAC-----TCTGATGATGACAAGAG-----4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
DB 4047 -----CCAAAGCGCAGAGGGCGTCCGAGAGGTGTCCGGAAGACCTCTGCGAGGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
DB 4098 TTTACTGATGCAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAGATT-----4145
QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTyrGlyLysIle 296
DB 4146 -----GGTCTCCCTCTTGAAACGGGTGGAGTCTTAGACAGCTGAT---GCTGAGCTG 4193

CC antisense nucleic acid molecule, a polypeptide or an antibody. The
 CC invention further comprises: an antisense nucleic acid sequence chosen
 CC from any of the sequences provided in the specification; preparation of
 CC the antibody comprising in vivo or in vitro immunisation of an
 CC immunocompetent animal cell, preferably of a vertebrate and most
 CC preferably of a mammal, with at least one of the polypeptide sequences
 CC chosen from a sequence provided in the specification; a mammalian
 CC expression vector comprising at least one antisense sequence chosen from
 CC an antisense nucleic acid provided in the specification; preparation of a
 CC genetically modified cell, that over- or under-expresses a gene
 CC implicated in an angiogenic disorder, comprising inserting the vector
 CC from above into a mammalian cell; a genetically modified cell that over-
 CC expresses or under-expresses at least one gene involved in angiogenesis
 CC by a nucleic acid sequence chosen from any of ADP73049 to ADP73338, as in
 CC the antisense nucleic acid selection of above, or a fragment of any of
 CC these; preparation of a cell line that stably expresses an expression
 CC vector. The angiogenesis inhibitor has the following activities:
 CC antiangiogenic, angiogenesis stimulator, angiogenesis inhibitor,
 CC cytostatic, dermatological, antiarthritic, antirheumatic,
 CC antiinflammatory, vasotropic, hypotensive, ophthalmological,
 CC antipsoriatic, and cardiant. The novel nucleic acid molecules of the
 CC invention may be used to treat disorders in gene therapy and antisense
 CC gene therapy. The nucleic acid sequences, proteins and antibodies as part
 CC of the therapeutic compositions are useful in treating a disorder of
 CC angiogenesis chosen from: tumour vascularisation, retinopathies,
 CC rheumatoid arthritis, Crohn's disease, atherosclerosis, ovarian
 CC hyperstimulation, psoriasis, endometriosis associated with
 CC neovascularisation, restenosis due to balloon angioplasty, tissue
 CC granulation due to scarification, peripheral vascular disorders,
 CC hypertension, vascular inflammation, Raynaud's disease, aneurism,
 CC arterial restenosis, thrombophlebitis, lymphadenopathy, lymphedema,
 CC ischaemia, angina, myocardial infarction, chronic heart disease,
 CC congestive heart disease, macular degeneration linked to age and
 CC osteoporosis. This polynucleotide represents a human DNA sequence for the
 CC creation of an angiogenesis inhibitor of the invention.

XX Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000127 Length: 7764
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conservative: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: 12 Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x ADP73081 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
 Db 3342 CGGGACACAGTCGTCTATCTTGAATCTGACTGGACCCCGACAGATGACTGCGAGGCACAA 3401
 QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 Db 3402 GCCCGAGCGCATAGATTCGTTCAGAACAGACAGGTAAATATTACCGCTAGTTACAAAG 3461
 QY 42 CysThrValGluGluLysAla-----Leu 49
 Db 3462 GGGACTGTGGAGAGAGATCATAGAACGGGCCAAAGAGAGTGTATTAGATCATCTGTG 3521
 QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
 Db 3522 GTGATTCAGCGCATGCACACCACTGCGCGACGATCTCTGMAAAACAACCTCAGGAAGGTCC 3581
 QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 Db 3582 AACTCAAATCTCTTTTAAATAAGAAAGAGCTGACACGCTATTTTGAAATTTGGAGCAGAGAT 3641
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 Db 3642 CTCCTCAAGAACTGAA-----GGGAGGAACTCA 3671
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120

Db 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTCGGTGGCTGAA 3716
 QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 Db 3717 ACGAGAGAGATGAAGTGTCAACAGTCCACAGATGAATCTCTA---TCACAGTTTAAG 3773
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer 160
 Db 3774 GTTGCCAACTTTGCAACAATGAAGATGAAGAAGACTA-----GAAGAGCGT 3821
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 Db 3822 CCTCACAAGGACTGGGATGAGATCAT-----3848
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 Db 3849 -----CCAGAGAAACAAAGAAAGTAAGTAGGAGGAAGAGCGGCAGAGAGCTAGNA 3902
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 Db 3903 GAAATTTATATGCTGCCTCGAATTCGGAGTTCCTCAATAAAGGCTCAGACAAATGACAGT 3962
 QY 204 AspGlyGlyAsnAlaLysArgLysLysLysAlaSerAspAspValThrAspProArgVal 223
 Db 3963 GACTCTGACACTGAGTCTTAAGAGGCGAGCCAGAGATCCTCTGCTTCGAGAGTGAACG 4022
 QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 Db 4023 GAAGAC-----TCGTATGATGACAAAG-----4046
 QY 244 LeuGluSerProLysValIleThrLeuGlnSerCysLysSerSerGlyThrAspGly 263
 Db 4047 -----CCAAAGCGCAGAGGGCGTCCGAGGAGTGTGCGAAGGACCTCTGCGAGGA 4097
 QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 Db 4098 TTTACTGTGACAGATCCGAAGTTCATCAAGGCTTATTAAGAAGTTT-----4145
 QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 Db 4146 -----GGTCTCCCTCTTGAACCGGTGGAGTGTGACACAGTGTAT---CCTGAGCTG 4193
 QY 297 ProAspGluSerGlnArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 4194 GTAGATAAGTCGTGGCAGATCTG-----AAGCGCCTGGGT 4229
 QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 Db 4230 GAACGTGATCCAC-----AACAGCTGTGTCTCAGCAATGCAGGAATATGAAGAGCAGCTG 4283
 QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
 Db 4284 AAAGAATATGCCAGCGAGGAAAGACCGAGGAAAGG---AGAGGTCCAACATCAAG 4340
 QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 Db 4341 ATATCCGGAGTT---CAGGTTAATGTGAATCCATTATTCACATCAAGAGAGGAGTTT---4394
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 Db 4395 -----GAGATGCTGCAATAATCTATCCTCTGTGGACCCCTGAGAGAAAAA 4442
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 Db 4443 TACTGCTTAACCTGTCGTGTCACAACTGCACATTTTGTAGTAGTGGGGTGGAGAT 4502
 QY 415 AspCys-----PheGly-ThrAsnSer-----ly 422
 Db 4503 GATTCCTCGCCTGTCGTGGGATTTTGAACATGCTGTAAGTGGAGTGGAGTAAATAA 4562
 QY 422 sGlnSerValSerThrLysLeuValAsn-----432

Db 4563 ACAG-----ACCCAGAGCTTAATTAACACAAATTCGCCGGTGGAGACA 4610
Qy 433 ----GluSerIeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db 4611 GATATAAAGGCTTCAGGGGAAGCAGCTACAGACCGGCGGATTAATCTGTTGTAAGCTGCTC 4670
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4671 AGAAAGGCTGGAGAGAAGAGGGGGTGTGCACAGGTGGGAGGAGGCCAAATTAAGAAG 4730
Qy 464 Arg----- 464
Db 4731 CGGAAGCTCGGTAAAGAGGAAACAAAGTGCAGGCTGAAAGAGGAGCATGGNATT 4790
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4791 GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAGGAGGAGAGATGAAGATGAT 4850
Qy 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4851 GGCTTGGAAAAAGTCCAAATGNAAAAAACAGAG---AAGAAAGAGAACAGGAGAAC 4907
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4908 AAGAGAAACAATAGTCTTAGGAAGACAAAGAGGGGACAGAGGAAAGAAAGTCA 4967
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4968 AAGATAAAGAAAGAGAGCGCTAAAGTGGTGATGCCAAATCTTCGAGTAATCAAGCGA 5027
Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 5028 TCTCAGGGTCTCTGTCATATTACAGCAGGAGTGAACCTCTGCCCATTTGGAGAGGATGAG 5087
Qy 539 ----- 539
Db 5088 GATGATGATCTGGACCGAGAGACATTCAGCATATGTAAAGAGAGATGAGGCCCGGTGAAA 5147
Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5148 AAGGCACCTGAACAGCTCGACAAACCTGACAGGGGCTCAACGTGCAAGAACAGCTGGAA 5207
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5208 CACACCGGAACTGCCTGCTGAAATCGGAGACCGGATAGCGAGTGCCTTAAAGCCTAC 5267
Qy 564 ----- 564
Db 5268 TCAGATCAGAGCAGACATCAAACTCTGGAGGAGAACCTATGGATTTTGTTCACAGTTT 5327
Qy 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 5328 ACAGATTTTGATCTCGAAACTGCATAGTTATACAGATGCTCATAGAAGAGTCT 5387
Qy 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaLysLeuLysValCys 597
Db 5388 CAAGAAGAGAG-----CAGCAAAAGAGAAAGAACGAC 5420
Qy 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5421 GTG-----ACTGGGGGTAGAAACCATTTTCCTCAGAGGCGCTCAGCGCGG 5471
Qy 614 -----IleSerGlnAsn 617
Db 5472 GACTCTCTGATATCTCAGTCC 5492

RESULT 14

ADS48467

ID ADS48467 standard; cDNA; 3045 BP.

XX

AC ADS48467;

XX

DT 02-DEC-2004 (first entry)

XX

DE Bacterial polynucleotide #3210.

XX

Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; call cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.

OS

Bacteria.

XX

PN US2003233675-A1.

XX

PD 18-DEC-2003.

XX

PF 20-FEB-2003; 2003US-00369493.

XX

PR 21-FEB-2002; 2002US-0360039P.

XX

(CAOY/) CAO Y.

PA

(HINK/) HINKLE G J.

PA

(SLAT/) SLATER S C.

PA

(CHEN/) CHEN X.

PA

(GOLD/) GOLDMAN B S.

XX

XX

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX

WPI; 2004-061375/06.

XX

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

XX

Claim 1; SEQ ID NO 26897; 122pp; English.

XX

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX

SQ Sequence 3045 BP; 1039 A; 712 C; 819 G; 475 T; 0 U; 0 Other;

Alignment Scores:

Align. No.:	5e-05	Length:	3045
Score:	164.00	Matches:	131
Percent Similarity:	37.16%	Conservative:	112
Best Local Similarity:	20.03%	Mismatches:	236
Query Match:	5.10%	Indels:	176
DB:	13	Gaps:	29

SEQ3_k705_COPY_650_1270 (1-621) x ADS48467 (1-3045)

```
QY 17 AspValLysHisValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePhe 36
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 673 GACATCGACATTTGGCGGCAAGTGCACACCGATCTCAAAAAGAGAGGAGGATTCTCTATCT 732
QY 37 ArgLeuTyrSerValCysThrValGluGluLysAlaLeuLysAlaLeuLysAlaLeuLys 56
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 733 CGCTTG-----CAGGAAGAGCT-----ACTGCAAAACCGAG 765
QY 57 ArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyr 76
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 766 CAGCGCCACCATGTCGGAATCTTAACCGACCGCTGAAC----- 807
QY 77 GlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSer 96
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 808 -----GATATCATGGATACCGCGCAGCGCTTGGCGTTCCTATCAAA 849
QY 97 GlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeu 116
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 850 GCGCGCGCGACCTCATGACATAGCGCAACATCTTGAAGTCTTCGATAACATATTG 909
QY 117 SerSerLysGlyGlyGluGlnAsnGluValLysLeuCysLeuLeuGluAlaLys--- 135
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 910 AACCTGTAC-----GAAGACACCGAGCAAGCTTGTTCAAACCGAAGCCAGGAC 963
QY 136 HisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGluAspHisIle---Lys 154
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 964 AAGGCAGACGACGCGCCAGGAATCCGCCACACGATCCGCGAGTGAACATCCAGAAG 1023
QY 155 LeuSerAspGluGluSer-----ProAsn 162
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1024 CTCACGCTCTGAACAAACAAAGCTTAAAGAACAGAGAGATGAGCAAAAAGAGCGCGAC 1083
QY 163 IlePheTyrSerLysLeuGlyGlyLysAsnProMetTyrLysTyrProSerAspThr 182
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1084 GCGTTTTACAGAACTC-----AGAAAGAGGATGGTTGAGCGCGCGAGGAT 1134
QY 183 Pro-----GlnArgAsnArgLysArgValGlnTyrPheGluGly 195
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1135 CCATCGCGCTCACTTCCCTCGACCAAGAGTCTTGAAGACCAACAGATCGCTGGGAG 1194
QY 196 SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysIleArg---LysLysAla 214
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1195 TCTGAAAGAAATGCCCTGAAACCAAGAAATCACTAATCTCAACAAATTCGATCGAAGGCC 1254
QY 215 SerAspAspValThrAspProArgVal-----ThrAspProProValAspAsp 230
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1255 AATGTGCGCTCAAGGACATGAAGCTTATGCAAGTGGAGGAGGATCTCTGTGCTGATCTA 1314
QY 231 AspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----ProLys 248
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1315 CAACACGTCAGGAA-----GGGATATCAAAAGCTTGAACCAACACAG 1356
QY 249 ValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsn 268
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1357 GTTCTGGACCTCAAGCAAAATTTGGCGAGCAAGCGCCGCGCGCGCA-----TCT 1409
QY 269 AspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeu 288
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1410 GAAGGATATTCATGCTTTTGGATTTGGAGAACAAAGTCTTTGGAGACAGCGCAACGCTGT 1469
QY 289 AlaSerGlnAspTrp-----GlyLys 295
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1470 GCAACCGAAAAGAGGATCTTTCAGCAGCAGCAGCTTTCTCAGACGCAAGCGAGGTGGA 1529
QY 296 IleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu 315
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1530 GGTGGAGACGACCTCAAGAA-----CAGCGGCTCGCAAAAGATGGATTT----- 1574
QY 316 CysGlnValLeuHisLeuSerAlaCysThrSerMetValGlyAsnPheLeuGluTyr 335
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 1575 -----GGAGACTCAAGAAATGA 1592
```

RESULT 15

```
AAQ10378
ID AAQ10378 standard; cDNA; 5858 BP.
XX
AC AAQ10378;
XX
DT 25-MAR-2003 (revised)
DT 12-APR-1991 (first entry)
XX
DE Encodes human 160kD mediator of inflammation protein.
XX
KW Mediator of Inflammation; cytokine; Hodgkin's lymphoma; MRP-160; ss.
XX
```

OS Homo sapiens.
 XX Key Location/Qualifiers
 FH 134..4417
 FT CDS
 FT /*tag= a
 FT /product= "MRP-160"
 FT /note= "mediator of inflammation"
 XX

XX EP412050-A.

XX PN

XX PD

XX 06-FEB-1991.

XX 26-JUN-1990; 90EP-00810481.

XX PF

XX 05-JUL-1989; 89GB-00015414.

XX PR

XX (CIBA) CIBA GEIGY AG.

XX PA

XX Odink KG, Tarsay L, Bruggen J, Wiesendang W, Cerletti N, Sorg C;

XX PI Dewolfpeet C;

XX DR WPI; 1991-038913/06.

XX DR P-PSDB; AAR10534.

XX XX

XX 160 kD human polypeptide mediator or precursor of inflammation -

XX PT polyclonal or monoclonal antibodies to polypeptide treat and diagnose

XX PT chronic inflammation and Hodgkins lymphoma.

XX XX

XX Claim 14; Page 32; 47pp; English.

XX CC mRNA isolated from human L132 cells was used to synthesize ds cDNA which

XX CC was cloned in phage Lambda gt10. The library was screened for the

XX CC sequence encoding the molecule which reacts with MAb1 C5 (see EP-

XX CC 162812), by successive screening with MAb1 C5 and cDNA inserts from the

XX CC commercially available cDNA libraries ML1005B and HL029B. One of the

XX CC positive clones contained a 6kb cDNA fragment which was found to encode

XX CC MRP-160. The sequence was cloned in pUC19 to give a hybrid vector which

XX CC was used to transform E.coli DH5-alpha cells. A plasmid yielding the

XX CC correct restriction fragments was designated pMRP160. (Updated on 25-MAR-

XX CC 2003 to correct PI field.)

XX SQ

Sequence 5858 BP; 1962 A; 1171 C; 1374 G; 1351 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.000163 Length: 5858
 Score: 163.00 Matches: 136
 Percent Similarity: 40.13% Conservative: 108
 Best Local Similarity: 22.37% Mismatches: 210
 Query Match: 5.07% Indels: 155
 DB: 2 Gaps: 28

SEQ3_K705_COPY_650_1270 (1-621) x AAQ10378 (1-5858)

QY 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
 DB 1841 GAAGAAAGCATCATCAGAAAGGAGATTAAGGCTCTGTATACCGCCACGGAAAG-----1891
 QY 51 LeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
 DB 1892 CTTTCCAAAGAGAACGAGTCATTGAAGCAAGCAAGCTGGAGCATCCCAACAAAGAGAACTCA 1951
 QY 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
 DB 1952 GATGTGATAGCTCTATGGAAGTCC-----AACTG-----1981
 QY 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
 DB 1982 -----GAGATGCTCATCGCATCCACAGCAGCGCATGGAAGAACTGGAAGGTA 2029
 QY 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLysLeu-----128
 DB 2030 TCCTTCAGCAAGGGCTTGGACACGAGACGGCGAGATTTGCTGNACTTAAACACAAATA 2089

QY 129 ---CysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSer-AspSerThrLe 147
 DB 2090 GAGAAATAGACTAGTATACCAACAGAAATAGAAATTTTGAGAAAGCAAGACTCT 2149
 QY 147 uPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSer-- 166
 DB 2150 GAACGGGCTGCCCATCTTAAAGAGA-----TGGAAAGCC 2182
 QY 167 -----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrPro 180
 DB 2183 TTGAGGGCTAAACTGATGAAGTTATTAAGAAAGAGAAACAGTCTGGAAGCCATCAGG 2242
 QY 180 erAspThrProGln-----ArgAsnArgLysArgValGlnTyrP 193
 DB 2243 TCGAAACTGGACAAAGCAGAACCCAGCATCTCGTAGAAATGGAAGACACGTTAAACAA 2302
 QY 193 heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys 210
 DB 2303 TTACAGAAAGCTGAAATAAAGGTAAAGAGCTAGAGGTACTGCAAGCCAAATGCAATGAA 2362
 QY 211 ArgLysLysAlaSerAspValThrAspPro---ArgValThrAspProProValAsp 229
 DB 2363 CAAACCAAGGTTAATGATAATTTTACATCACAGCTCAAGGCTACTGAAGAAAGCTCTTG 2422
 QY 230 AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu 245
 DB 2423 GATCTTGATGCACTTCGAAAGCCAGTTCGGAAGTAAATCGGAAATGGAAGAACTTAGA 2482
 QY 246 SerPro-----LysValIleThrLeuGln-----Ser 254
 DB 2483 CAGCAGCTTGAGGCAGCTCAGAAACAGATTAAACATTTAGAGATTGAAAGAATGCTGAA 2542
 QY 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
 DB 2543 AGTAGCAAGCTAGTAGCAATACCAGAGAGCTCCAGGGGAGAGAG---CTAAAGCTTACT 2599
 QY 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly 294
 DB 2600 AACCTTCAGAAATTTGAGT-----GAAGT-CAGTCAAGTGAAGAAAGCTTTGGA 2649
 QY 295 LysIleProAspGluSerGlnArgArgLeu-----HisThrValLeu 308
 DB 2650 AAAGAACTTCAGATTTTGAAGAAAGTTTGTCTGAAGCTTCAGAGAGGAGCGTCTCTGT 2709
 QY 309 LysProLysMetAlaLysLeuCysGln-ValLeuHisLeuSerAspAlaCysThrSerMe 328
 DB 2710 TCAGAGAAGTATGCAAGAACTGTAATAAGTTACACCAAGAGGAGGAGACAGTTAAACAT 2769
 QY 328 tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaTh 348
 DB 2770 GCTGTCTCTGACTTGGAGAGCTGAGAGAAACTTAGCAGATATGAGGCAAAA-----2824
 QY 348 rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuValLysGlnIle 368
 DB 2825 -----TTTAGAGAGAA 2835
 QY 368 eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
 DB 2836 AGATGAGAGAGAGAGAGAGCTGTAAGGCAAGGAAAGAACTG-----2878
 QY 388 gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi 408
 DB 2879 ----GAAATGAC-----ATTGCAAGAAATAATGAAG-----2905
 QY 408 sthrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerth 428
 DB 2906 -----ATGTGAGAGATTAATCTTCTCAG-----CTGAC 2934
 QY 428 rLysLeuValAsnGlu---SerLeuSerGlyAlaThrValArgAspGluLysIleAsn-- 446
 DB 2935 AAAGATGAACGATGAATTAACGTCTGNAAGAAAGAGATGTAGAAGAAATTACAGTAAACT 2994
 QY 447 -ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrThrGluLysArgCysSe 466


```
Db 2760 GCGGACACAGTCGTCACTCTTGAAGCTGAGTCCGAAACCCAGAGTGAAGTTCGAGGACAA 2819
Qy 22 GluLysIleLeuLysIleGluSerCysSerGluArgThrLysIlePheArgLeuThrVal 41
Db 2820 GCCCGAGCCCATAGATTGGTCAGAGAAGACGAGTAAATATTACCGCTTAGTTACAAG 2879
Qy 42 CysThrValGluGluLysAla
Db 2880 GGGAGCTGTGGAGGAGAGATCATAGAACGGGGCCMAAAGAGAGATGGTATTAGATCATCTG 2939
Qy 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn
Db 2940 GTGATTACGCGATGACACCACTGCGCGGACGATCTCGTGAACAACTCAGGAAGTCC 2999
Qy 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3000 AACTCAATCCCTTTTAAAGAGAGCTGACAGCTATTATTGAATTTGAGCAGGAGAT 3059
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3060 CTCTTCAAGAACTGGAA-----GGGGAGGAATCA 3089
Qy 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly 120
Db 3090 GAACCTCAGAAATGGAT-----ATAGATGAATTTTGGCGTTGGCTGAA 3134
Qy 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3135 ACAGAGAGAGANTGAAGTGTCAACAGTGCACAGATGAATCTTA---TCACAGTTTAA 3191
Qy 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLeuLeuSerAspGluGluSer 160
Db 3192 GTTGCCAACTTTGCAACAATGGAAGATGAAGAAGAGCTA-----GAAGAGCGT 3239
Qy 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
Db 3240 CCTCAAGGACTGGGATCAGATCAAT----- 3266
Qy 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu
Db 3267 -----CCAGAGGACAAAGGAAAGAAAGTAGAGGAGAGAGCGGAGAGGAGGTAGAA 3320
Qy 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3321 GAAATTTATATGCTGCTCGAATTCGAGTTCCACTAAAGAGGCTCAGACAAATGACAGT 3380
Qy 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db 3381 GACTCTGACACTGAGTCTTAAGAGGCGAGCCGAGAGATCCTCTGCTTCTGAGAGTGAACG 3440
Qy 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 3441 GAAGAC-----TCTGATGATGACAAGAG----- 3464
Qy 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 3465 -----CCAAAGCGCAGAGGCGCTCCGAGAGTGTGCGAAGGACCTCGTGGAGGGA 3515
Qy 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
Db 3516 TTACTGATGCGAGAGATCCGAAGGTTTCATCAAGGCTTATAAGAGTTT----- 3563
Qy 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db 3564 -----GGTCTCCTCTTGAAACCGCTGGAGTGCATAGCAGCTGAT---GCTGAGCTG 3611
Qy 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 3612 GTAGATAAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 3647
Qy 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGlyThrVal 336
Db 3648 GAACCTGATCCAC-----AACAGCTGTGTCTCAGCAATGCAGGAATACGAAGACGAGCTG 3701
```

```
Qy 337 IleGluAsnHisArgIleTyrGluGluProLThrThrPheGlnAlaPheGlnIleAla 356
Db 3702 AAAGAAATATCCAGCGAGGAAAGACGACGAGGAAAGG---AGAGTCCCAACAATCAAG 3758
Qy 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
Db 3759 ATATCCGAGTT---CAGGTTAATGTGAATATCCATTATCCAAACATGAAGAGGAGTTT--- 3812
Qy 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
Db 3813 -----GAGATGCTGCATAAATCTATCCCTGTGGACCTCGAAGAAAAA 3860
Qy 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 3861 TACTCTTAACCTGTCGTCAAGCTGCACATTTTGTATGTAGAGTGGGGGTGAAGAT 3920
Qy 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
Db 3921 GATTCCTCCCTGTTGCTGGGATTTATGAACATGGCTATGGAACACTGGGAGTTAATTAA 3980
Qy 422 sGlnSerValSerThrLysLeuValAsn----- 432
Db 3981 ACAG-----ACCAGAGCTTAAATTAATCTGACAAAATTTCTCCCGTGGAGACA 4028
Qy 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
Db 4029 GATAAAAGCCTCAGGGGAGCAGCTACAGACCCGAGGGATTACTTTGTAAGCTGCTC 4088
Qy 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4089 AGAAAGGCTGCGAAGAAAGGGGCTGTGACAGGTGGGGAAGGCCAAATTAAGAAG 4148
Qy 464 Arg----- 464
Db 4149 CGGAAGCTCGGTTAAAGAGAAACAAAGTCCAGCTCAAGAGGAGCATGGAAAT 4208
Qy 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrLeSer 480
Db 4209 GAGCTTTTCATCTCCTAGGCAATTCAGATAATCCATCAGAAGAGGAGAGATGAAGATGAT 4268
Qy 481 GlyIleLysLysLysLysValGlnLysLeuValGlnLysLeuValGlnLysHisGluGluLys 500
Db 4269 GCGTTGGAAAAAGTCCAAATGAAAAAACAAGAG---AAGAAAGAGAACGAAGAAC 4325
Qy 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4326 AAGGAGAAACAATGAGTTCTAGGAAGACAAAGAAAGGGGCAACAGAAAGAAAGTCA 4385
Qy 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4386 AAAGATAAGAAAGAGAGAGCTAAAAAGTGTGTGATGCTGCAAAATCTTCGAGTAAATCAAAGCGA 4445
Qy 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 4446 TCTCAGGCTCTGTCCATATTACAGCAGAGAGTGAACCTGTCCCATTTGGAGAGGATGAG 4505
Qy 539 ----- 539
Db 4506 GATGATGATCTGGACCAGGAGACATTCAGCATATGTAGAGGAGGATGAGGCCCGTGA 4565
Qy 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 4566 AAGGCACCTGAAACAGCTCGACAAACCTGCAAGGGGCTCAACGCTCAACAGCAAGCTGGAA 4625
Qy 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 4626 CACACCCGGAATCGCTGCTGTAATAATCGGAGACCGGATAGCCGAGTGCCTTAAGCCTAC 4685
Qy 564 ----- 564
Db 4686 TCAGATCAGGAGCACATCAAACTCTGGAGGAGGAACCTATGATTTTGTTCCTCAAGTTT 4745
```

```

QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 4746 ACAGAAATTGATGCTCGAAACTGCATAGCTTATACAGAGTGCCTCATAGAAAAGGTCT 4805
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 4806 CAGAGAGAGAG-----GAGCAAAAGAGAAAGAGCAGC 4838
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 4839 GTG-----ACTGGGGTAAGAAACCAATTTTCGTCAGAGGCTTCAGGCTCCAGCGCG 4889
QY 614 -----IleSerGlnAsn 617
Db 4890 GACTCTCTGATATCTCAGTCC 4910

RESULT 2
US-09-949-016-2989
; Sequence 2989, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2989
; LENGTH: 7699
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2989

Alignment Scores:
Pred. No.: 1,06e-07 Length: 7699
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 4 Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x US-09-949-016-2989 (1-7699)
QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
Db 3343 CGGACACACAGCTGCTATCTTGTACTGTGCTGAAACCCAGATGACTTGCAGGACAA 3402
QY 22 GluLysIleLysIleGluSerCysSerGluArgTrpLysIlePheArgLeuTrpSerVal 41
Db 3403 GCCCGAGCGCATAGATTGCTCAGAGAAGCAGGTAAATATTACCGCTTAGTTACAAAG 3462
QY 42 CysThrValGluGluLysAla-----Leu 49
Db 3463 GGGACTGTGGAGGAGGAGATCATAGAACGGGCCAAAGAGAGATGTTATGATCATCTG 3522
QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
Db 3523 GTGATTCAGCGCATGACACCACTGCGCGACACATCTCTGGAACAACTCAGAGAGTCC 3582
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTrp 80
Db 3583 AACTCAAAATCTTTTAAAGAAAGAGAGCTGACAGCTATTTTTGAATTTGGAGCAGAGGAT 3642
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100

```

```

Db 3643 CTCCTCAAGAACTGGAA-----GGGAGGAGATCA 3672
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 3673 GAACCTCAGGAAATGGAT-----ATAGATGAAATTTTTCGGTTGGCTGAA 3717
QY 121 GlyGluGlnAsnGluValLysLysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 3718 ACAGAGAGAGAAATGAAGTGTCAACAGTGCACAGATGAACCTTCTA---TCACAGTTTAA 3774
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
Db 3775 GTTGCCCACTTTGCCAACAATGGAGATGAAGAAGACTA-----GAAGAGCGT 3822
QY 161 ProAsnIlePheTrpSerLysLysLeuLeuGlyGlyLysAsnProMetTrpLysTrpProSer 180
Db 3823 CCTCACAAGGACTGGGATGAGATCAAT-----3849
QY 181 AspThrProGlnArgAsnArgLysArgValGlnTrpPheGlu-----194
Db 3850 -----CCAGAGGAACAAAGGAAAGTAGAGGAGGAGCGGAGAGAGTAGAA 3903
QY 195 -----GlySerGluAlaSerProLysThrGly 203
Db 3904 GAAATTTATATGCTGCTCGAATTCGAGTTCACCTAAAGAGCTCAGACAAATGACAGT 3963
QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
Db 3964 GACTCTGACACTGAGTCTAAGAGGCGCCAGAGATCTCTGCTTCTGAGAGTGAACG 4023
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
Db 4024 GAAGAC-----TCTGATGATGACAAGAA-----4047
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
Db 4048 -----CCAAAGCGCAGAGGCGTCCGAGGAGTGTGGGAGAGACCTCTGCGAGGGA 4098
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTrpSerMetGlySer 278
Db 4099 TTTACTGATGCAGAGATCCGAGGTTTCATCAAGGCTTATAAGAAAGTTT-----4146
QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
Db 4147 -----GGTCTCCCTCTTGAACGGCTGGAGTGCATAGCAGCTGAT---GCTGAGCTG 4194
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
Db 4195 GTAGATNAGTCGTGGCAGATCTG-----AAGCGCTGGGT 4230
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTrpVal 336
Db 4231 GAACGTATCCAC-----AACAGCTGTGTGTCAGCAATCAGGAATACGAGAGCAGCTG 4284
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
Db 4285 AAAGAAATGCCAGCGAGGAGGAAAGGACGAGGAAAGG---AGAGTCCCAACAATCAAG 4341
QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysLeuSerLeuVal 376
Db 4342 ATATCCGGAGTT---CAGGTTAATGTGAATTCATTATCAACATCAAGAGGAGTTT---4395
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTrpIle 394
Db 4396 -----GAGATGTCATAAATCTATCCCTGTGACCCCTCAAGAAAAA 4443
QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
Db 4444 TACTGCTTAACCTGTCGTGTCAAAGCTGCACATTTTGTATGATGAGTGGGGGTGGAAGAT 4503
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
Db 4504 GATTCTCGCTGTGTGGGGATTTTATGAACATGCGCTATGGAAACTGGGAGTTTAATTA 4563

```

```
QY 422 sGlnSerValSerThrLysLeuValAsn----- 432
Db 4564 ACAG-----ACCAGAGCTTAAATTAACTGACAAAATTCGCGGTGGAGACA 4611
QY 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
Db 4612 GATAAAAGCCTCAGGGGAAGCAGCTACAGCCGAGCGGATTTACTTTGTAAGCTGCTC 4671
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
Db 4672 AGAAGGGTCTCGAGAGAAGGGGGCTGTGACAGGTGGGAGAGAGCCAAATTAAGAAG 4731
QY 464 Arg----- 464
Db 4732 CGGAAGCCTCGGTAAAGAAAGAAACAAAGTCCAGCTGAAAGAGGAGCATGGAATT 4791
QY 465 -----CysSerHisThrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 4792 GAGCTTTTCATCTCTAGGCATTCAGATAATCCATCAGAAGAGGGGAGAGTGAAGATGAT 4851
QY 481 GlyIleLysLysLysThrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
Db 4852 GCCTTGAAGAAAGTCCATGATGAAGAAACACAGAG---AAGAAAGAGAACAGGAGAAC 4908
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
Db 4909 AAGGAGAAACAAATGAGTTCTAGGAAAGACAAAGAGGGGACAAAGGAAAGAAAGTCA 4968
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
Db 4969 AAGATAAAGAAAGAGAGCCTTAAAGTGGTGATGCCAAATCTTCGAGTAAATCAAAGCGA 5028
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp----- 539
Db 5029 TCTCAGGGTCTGTGCATATTACAGCAGGAAGTGAACCTGTCCCCATTGAGAGGATGAG 5088
QY 539 ----- 539
Db 5089 GATGATGATCTGGACAGGAGACATTCAGCATATGTAAGGAGAGGATGAGGCCGTGAAA 5148
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
Db 5149 AAGGCACCTGAAACAGCTGCACAAACCTGCAAGGGGCTCAACGTGCAAGAACAGCTGGA 5208
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
Db 5209 CACACCCGNACTGCTGCTGAAATCGGAGACCGGATAGCCGAGTGCCTTAAAGCCTAC 5268
QY 564 ----- 564
Db 5269 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGGATTTTGTTCCAAGTTT 5328
QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
Db 5329 ACAGAAATTTGATGCTCGAAACTGCATTAAGTTATACAGATGGCTCATAGAAAGAGTCT 5388
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
Db 5389 CAAGAGAAGAG-----GAGCAAAAGAAAGAACGAC 5421
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
Db 5422 GTG-----ACTGGGGTGAAGAAACCAATTCGTCCAGAGGCTCAGGCTCCAGCCGG 5472
QY 614 -----IleSerGlnAsn 617
Db 5473 GACTCTCTCATATCTCAGTCC 5493
```

RESULT 3

US-09-220-132-79

; Sequence 79, Application US/09220132

; Patent No. 6506607

```
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 5857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-79
```

```
Alignment Scores:
Pred. No.: 4,37e-07 Length: 5857
Score: 159.00 Matches: 136
Percent Similarity: 39.97% Conservative: 107
Best Local Similarity: 22.37% Mismatches: 211
Query Match: 4.94% Indels: 155
DB: 4 Gaps: 28
```

SEQ3_K705_COPY_650_1270 (1-621) x US-09-220-132-79 (1-5857)

```
QY 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 1840 GAAGAAACTCATCAGAAAGAGATAAAGCTCTGTATACCGCCACGGAAG----- 1890
QY 51 LeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 1891 CTTTCCAAGAGAACAGTCATTGAAAGCAAGCTGGAGCATGCCAACAAAGAGAACTCA 1950
QY 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
Db 1951 GATGTGATAGTCTCTATGGAAGTCC-----AAACTG----- 1980
QY 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
Db 1981 -----GAGACTGCCATCGCATCCACACGAGAGCGCATGGAAGAACTGAAGGTA 2028
QY 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLysLeu----- 128
Db 2029 TCTTTCAGAAAGGGCTTGGAAACAGACAGACGCGCAGAAATTTGCTGAACTAAACACAAATA 2088
QY 129 ---CysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSer-AspSerThrIle 147
Db 2089 GAGAAATGACATAGATTACCAACACGAAATAGAAATTTTGCAGAAATCAACAAGACTCT 2148
QY 147 pHeGlyGluAspHisIleLysLeuSerAspGluSerProAsnIlePheTrpSer-- 166
Db 2149 GAACGGGTGCCCTGCTAAAGAGA-----TGGAGGCC 2181
QY 167 -----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrProS 180
Db 2182 TTGAGGGCTAAACTGATGAAGTTATTAAAGAAAGGAAACACAGTCTGGAAGCCATCAGG 2241
QY 180 exAspThrProGln-----ArgAsnArgLysArgValGlnTyrP 193
Db 2242 TCGAAACTGGACAAACAGACAGAACCCAGCATCTCTGTAGAAATGGAAGACACGTTAAACAAA 2301
QY 193 heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys 210
Db 2302 TTACAGGAAGCTGAAATTAAGTAAGGAGTAGAGTACTCCAGCCCAATTCGATGAA 2361
QY 211 ArgLysLysAlaSerAspAspValThrAspPro---ArgValThrAspProValAsp 229
Db 2362 CAACCAAGGTTATTGATAATTTTACATCACAGCTCAAGGCTACTGAAAGAAAAAGCTCTTG 2421
```


Qy	230	AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu	245
Db	2422	GATCTTGATGCACTTCGGAAGCCAGTTCCGAAGCTAAATCGGAATGGAAGAACTTAGA	2481
Qy	246	SerPro-----LysValIleThrLeuGln-----Ser	254
Db	2482	CAGCAGCTTGAGGCAGCTGAGAAACAGATTAAACATTTAGAGATTGAAAGAAATGCTGAA	2541
Qy	255	SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr	274
Db	2542	AGTAGCAAGGCTAGTAGCATTTACAGAGAGCTCCAGGGAGAGAG---CTAAGCTTACT	2598
Qy	275	SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly	294
Db	2599	AACCTTCAGGAAAAATTTGAGT-----GAACT-CAGTCAAGTGAAGAGAGACTTTGGA	2648
Qy	295	LysIleProAspGluSerGlnArgArgLeu-----HisThrValLeu	308
Db	2649	AAAAGAACCTTCAGATTGTAAGAAAGAAAGTTTGCTGAAGCTTCAGAGGAGCGAGCTCTGT	2708
Qy	309	LysProLysMetAlaLysLeuCysGln-ValLeuHisLeuSerAspAlaCysThrSerMe	328
Db	2709	TCAGAGAAGTAGTCAGAGAACTGTAATAAGTTACACCAAGAGAGAAACAGTTTAACAT	2768
Qy	328	tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaTh	348
Db	2769	GCTGCTCTCTGACTTGCAGAGAGCTGAGAGAAAACTTAGCAGATATGCAGGCAAAA	2823
Qy	348	rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeuValLysGlnIle	368
Db	2824	-----TTTAGAGAGAA	2834
Qy	368	eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr	388
Db	2835	AGATGAGAGAGAGAGCAGCTGTATAAAGCCAAAGGAAAAAATG	2877
Qy	388	gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi	408
Db	2878	-----GAAAATGCAC-----ATTGCAGAAATAATGAAG-----	2904
Qy	408	sThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValSerTh	428
Db	2905	-----ATGTCAGGAGATAACTCTTCTCAG-----CTGAC	2933
Qy	428	rLysLeuValAsnGlu---SerLeuSerGlyAlaThrValArgAspGluLysIleAsn--	446
Db	2934	AAATAATGAACGATGAATACGCTCGAAGAAAGAGAGATGTAGAAATTCACGCTAAACT	2993
Qy	447	-ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSe	466
Db	2994	TACAAAGGCTAAATGAAAATGCAAGT-----	3018
Qy	466	rHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysIleTy	486
Db	3019	-----TTTCTGCAAAAAGATTAATTGAGGACATGCTGTCAAAAGC	3056
Qy	486	rLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAs	506
Db	3057	TGAACAGCAGCCAGCAGACAGCCTAAAAAGCATGAGAAAGAAAGAAAGAAATTTGAGAG	3116
Qy	506	nMetTyrAlaAspLysLysGlnLysLeuGluThrSer-----LysSerValG1	522
Db	3117	GAAATTCTCGACCTGGAAGAAAGAAATGGAACCAAGCCACACAGCTGTACAGGAGCTGAA	3176
Qy	522	uAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu----	540
Db	3177	AGCCAGGTATGAGAGAGCCACT---TCTGAGACAAAAACCAAGCATGAAGAAATCTCTACA	3233
Qy	541	-----LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSe	556
Db	3234	GAACCTCCAGAGACGCTGCTGGAC-----ACAGAGCAAGCTGAAGGGCGCAGCGGA	3287
Qy	556	rgLuLysAsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLe	576

```

Db      3288 GGAGAACAGTGGCTTGCTGCAGGAGCTGGAGAGCTGAGAAAACGAAGCCGCAAAAGCCA 3347
           |||   ::   |||||:::   ::   |||   |||   |||
Qy      576 ualaGluaSpGuLa 581
           |||   |||   |||
Db      3348 AGTCGTCAAACAGCG 3363
           |||   |||   |||

RESULT 4
US-09-949-016-5001
; Sequence 5001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5001
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5001

Alignment Scores:
Pred. No.:      8,68e-07      Length:      5883
Score:          156.50        Matches:     133
Percent Similarity: 35.90%    Conservative: 114
Best Local Similarity: 19.33% Mismatches:    236
Query Match:       4.87%     Indels:       207
DB:                4         Gaps:        26

SEQ3_K705_COPY_650_1270 (1-621) x US-09-949-016-5001 (1-5883)

Qy      22 GluYslleIlylleGIuSerCysSerGIuArgThrIlyllePheArgLeuTyrSerVal 41
           ::::|||||:::|||||:::|||||:::|||||
Db      2878 CAGAAGCTGCAGCTGGAGAGGTGCACCACCGAGGCGAAGCTGAAAAAGCTG----- 2928
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Qy      42 CysThrValGIuGIuLysAlaLeuIIleLeuAlaArgGlnAsnLysArgGlnAsnLysAla 61
           |||||:::   :::|||||:::|||||:::|||||
Db      2929 -----GAGGAGGAGCAGATCATCTCTGGAGGACCAGAACTGCAAGCTGCCCAAG--- 2976
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Qy      62 ValGIuAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81
           |||||:::|||||:::|||||:::|||||
Db      2976 ----- 2976

Qy      82 PheAspLysLeuAspHisPheHisSerSerGIuThrProAspSerGlyValSerPheGlu 101
           |||||:::|||||:::|||||:::|||||
Db      2977 -----GAA 2979
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Qy      102 GlnSerIleMetAspGlyValIIleHicGIuPheSerSerIleLeuSerSerLysGlyGly 121
           ::::|||||:::|||||:::|||||:::|||||
Db      2980 AAGAAACTGTCTGGAAGACAGAATAAGCTGAGTTCCACCACCACTCACAGAA----- 3030
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Qy      122 GluGIuAsnGIuValLysLeuCysLeuLeuLeuGIuAlaLysHisAlaGlnGlyThrTyr 141
           |||||:::|||||:::|||||:::|||||
Db      3031 GAGGAGGAGAATCTTAAGAGCTCGCCNAGCTCAAGAACAAAGCAT----- 3075
           |||||:::|||||:::|||||:::|||||
Qy      142 SerSerAspSerThrLeuPheGlyGIuAspHisIleLysLeuSerSerAspGIuGlu-SerPr 161
           |||||:::|||||:::|||||:::|||||
Db      3076 -----GAGGCAATCATCACTGACTTGGAAAGAGCGCTCCCGAGGAGGAGAGAGCAG 3126
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Qy      161 oAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAs 181
           ::::|||||:::|||||:::|||||
Db      3127 CGACAGAGCTGGAGAAGACCGCCGGAAGC-----TGGAGG-----GA 3165
           |||   |||   |||   |||   |||   |||   |||   |||   |||   |||

```


Db 2938 AAGAAGGAACTGGAGGATGATGTTTCAGAGCTCAAGAAA----- 2979
Qy 41 ValCysThrValGluGluYsAlaLeuIleLeuAlaArgGlnAenYsArgGlnAenLys 60
Db 2980 -----GACATTGATGACCTTGAGCTGACACTGGCCAAAGTTGAGAGGAGAAACAT--- 3030
Qy 61 AlaValGluAenLeuAenArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
Db 3031 GCCACAGAGAACAGGTGAAACCTCACAAGAG-----ATGGCAGGTCTGGATGAA 3084
Qy 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
Db 3085 ACCATTGCTAAGCTG-----ACCAAGAGAGAGAGCTCTCCAGGAGGCCAC 3132
Qy 101 GluGlnSerIleMetAspGlyValIleHisGlu-----PheSerIleLeuSerSer 118
Db 3133 CAGCAGACCTGATGACCTGCAGATGGAGGAGGACAAAGTCAACCTTCACCAAGCT 3192
Qy 119 Lys-----GlyGluGluAenGluValLys 127
Db 3193 ANAACCAAGCTAGACAGCAAGTGGACGATCTTGAAGGATCTCTGGAAACAAGAAAGAAA 3252
Qy 128 LeuCysLeuLeuLeuGlu-----AlaLysHisAlaGlnGly-----ThrTyrSerSer 143
Db 3253 CTTTGGATGACCTTAGAAGACGACAGAGAAACTGGAGGTGACCTAAATAATTGGCCCAA 3312
Qy 144 AspSerThrLeuPheGlyGlu-----AspHisIleLysLeuSerAspGluGluSerProAsn 162
Db 3313 GAATCCACATGATACAGAAATGACAAACAGCAACTTAATGAGAACTCAAAAAGAAA 3372
Qy 163 IlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182
Db 3373 GAGTTTGAATGAGCAATCTGCAAGGCAAGATGAAAGATGAACCAAGCCCTTGCAATGCAG 3432
Qy 183 ProGlnArgAsnArgLysArgValGln---TyrPheGluGlySerGluAlaSerProLys 201
Db 3433 CTAACAAGAGATCAAGATCAAGATTAAGCCGCGGATGAGGAGCTGGAGGAGGAATCGAG 3492
Qy 202 ThrGlyAspGlyGlyAsnAlaLysLysArgLysAlaSerAsp----- 216
Db 3493 GCAGAGCGGCGCTCCCGGCCAAGCAGAAAGCAGCGCTCTGACCTCTCCCGGAGCTG 3552
Qy 217 ---AspValThrAsp-----ProArgValThrAspProProValAsp--- 229
Db 3553 GAGGAGATCAGTCAGAGGCTGGAAGAACCGGTGGGCCCACTTCAGCCCAAGATTGAGTTG 3612
Qy 230 AspAspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysVal 249
Db 3613 AACAAGCGGAGGCTGAGTTCCAGAAATCGCGGAGCTGGAA----- 3660
Qy 250 IleThrLeuGlnSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAsp 269
Db 3661 -----GAGTCCACCTGCAGCAGCAAGCCAGCCAGCTCTCTCGGAAGAGCAC 3711
Qy 270 AlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyPro----- 284
Db 3712 GCAGATAGTGTGCTGAGCTTGGGAAGCAGATCGACAGCTTCAGCGGCTCAAGCAGAG 3771
Qy 284 ----- 284
Db 3772 CTGGAGAGAAAGATGAGTGAAGATGGAGATCAATGACCTTGCTAGTAACTAGGAG 3831
Qy 285 -----GluAspMetLeu 288
Db 3832 ACTGTCTCCAAAGCCAAAGCAACTTTGAGAAATGTGGCCACCTTAGAGGACCACTT 3891
Qy 289 AlaSerGlnAspTrpGlyLysIleProAspGluSerGlnArgArgLeuHisThrValLeu 308
Db 3892 AGTGAATA-----AAAACAAGAGAGAGAGCAGCAGCAAGCTTAATAATGAGTTG 3942
Qy 309 LysProLysMetAlaLysLeu-----CysGlnValLeuHisLeu 321
Db 3943 TCAGCCCAAGGACGCTTTACACAGAAATCAGGTGAGTTTTCACGACGCTAGATGAA 4002

Qy 322 SerAspAlaCysThrSerMetValGlyAsn-----PheLeuGluTyrValIle 337
Db 4003 AAAGATGCTATGTTCTCAGCTATCCGAGGCAAAACAGCAATTTACACAACAGATTGAA 4062
Qy 338 GluAsnHisArgIleTyrGluGlu-----ProAlaThrThrPheGlnAlaPhe 353
Db 4063 GAATTAAGAGGAGGAGCTAGAGAGGAGACTAAGGCCAAGAGCACTCTGGCCCCATGCCCTG 4122
Qy 354 GlnIleAlaLeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGlu 373
Db 4123 CAGTCAGCC---CGCCATGACTGTGACCTGCTCGGGGAACAGTATGAGGAGGAGCAGAA 4179
Qy 374 Ser-----LeuValArgAlaAsnSerGluLeuAlaPheLysCys 386
Db 4180 GCCAAGGCTGAGCTGCAGAGGAGGAATGTCCAAGGCCAACAGTGAAGTGGCCATGGAGG 4239
Qy 387 SerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeu 406
Db 4240 ACCAAGTACGAGAGCAGCGCCATC----- 4263
Qy 407 GluHisThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValVal 426
Db 4264 CAGCGCACAGAGGAGCTGGAG-----GAG, 4287
Qy 427 SerThrLysLeuValAenGluSerSerGlyAlaThrValArgAspGluLysIleAsn 446
Db 4288 GCCAAGAAGAAGCTAGCCCGCGCTCTGCAGGATGCAGAAACATGTAGAAAGCTGTGAAT 4347
Qy 447 ThrLys-----SerMetArgAsnSerSerGluAspGluGluCys 459
Db 4348 TCCAAATGTGCTCTCTTGAANAAGACAAAGCAGAGGCTACAGAAATGAAGTAGAGGACCTC 4407
Qy 460 Met-----ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 4408 ATGATTGATGTGGAACGATCTAATGCTGCTGATAGCTCTCCATAGAAGCAAGAAAC 4467
Qy 475 IleGluLysThrIleSerGlyLysLysLysLysLysGlnValGlnLysLeuVal 494
Db 4468 TTTGACAAGGTTCTGCGCAGATGGAACAGAGTATGAGGAAACTCAGGCTGAACCTTGAG 4527
Qy 495 GlnGluHisGluGluLysLys-----MetGluLeuLeuAsnMetTyr 508
Db 4528 GCTCCCAAGAGGAGTCGCGTTCTCTCAGCAGCTGAGCTGTTCAGGTGAAGAAATGCTAC 4587
Qy 509 AlaAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIle 528
Db 4588 GAGGAATCCTGGATCATCTTGAACTCTAAGCGA---GAGNAATAGAAGCTTACACAG 4644
Qy 529 ThrCysSerArgThrSerThrGlnValGlyAspLeuLysLeuLeuAspHisAsnTyrGlu 548
Db 4645 GAGATTCTGACCTGACAGAGCAAAATGCGAGGGTGGAAAGCATATCCATGAACCTGGAG 4704
Qy 549 ArgLysPheAspGluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluMet 568
Db 4705 AAAGTAAAGAAACAACTTGTATCATGAGAGAGTGAACACAGACTTCCCTTAGAGGAACA 4764
Qy 569 -----HisGluValAlaLysLysLysLeuAlaGluAspGluAlaCysTrp 583
Db 4765 GAGGCAATCTCTGAGCATGAAGAGGCAAAATTTCTTCGCAATTCACCTTGAG----- 4815
Qy 584 IleAsnArgIleLysSer 589
Db 4816 CTAAATCAGGTGAATCT 4833

RESULT 6

US-09-949-016-2264

; Sequence 2264, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF


```
QY 382 euAlaPheLysCysSer-----ArgValGluVal-----AspTyrIleTyrS 396
D 3887 TGCAGTCCAGTCAGCGATGGGAGCGGCGCGGAGCTCAACGACGAAGTCCACA 3946
QY 396 erIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspC 416
D 3947 AGCTCGAGATGAAGTGGAGGCGTC----- 3972
QY 416 ysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLys 436
D 3973 --ACGGCATGCTCAGCGAGCGCGAGGGAAGGCCATCAAGCTGGCCACGAGGTGGCGT 4030
QY 436 er---GlyAlaThrValArgAsp-----GluLysI 445
D 4031 CCCTCGGGTCCAGCTCCAGGATACCCAGGAGCTGCTCCAAAGAAGAAACCCGCGCAGAAGC 4090
QY 445 leAsnThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgC 465
D 4091 TCAACGTGTCCACCAAGCTGGCAGCTGGGAGGACGAGAGAAACAGCTGCAGGAGCAGC 4150
QY 465 ysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIle----- 482
D 4151 TGCACGAGGAGATGGAGGCCAACAGACAACTGGAGCGCCACATCTCCACCCTGAACATCC 4210
QY 482 ----- 482
D 4211 AGCTCTCCGACTCAAAGAAAGAGCTGCAGGACTTTGCCAGCACCGTGGAGTCTTGGAGG 4270
QY 483 -----LysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysL 501
D 4271 AAGGCAAGAAGAGGTTCCAGAAGGAATTTGAGAGCCTCCACCCAGCAGTACGAAGAGAAA- 4329
QY 501 ysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSerLysSerV 521
D 4330 -----GCAGCTGCTTACGATTAACCTGGAAGAACCCAG- 4362
QY 521 alGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeuL 541
D 4363 -----AACAGGCTTCAGCAGGAGCTGGACGACCTGG 4393
QY 541 ysLeu---LeuAspHis-----AsnTyrGluArgLysP 551
D 4394 TCGTAGCTGGATATACCCAGCGCAACTGGTGTCACCTGGAAAAGACGACGAAGT 4453
QY 551 heAspGluIleLysSerGluLysAsnGluCysLysLeuLysSerLeuGluGlnMetHisGluV 571
D 4454 TCGATCAGTTGTTAGCGGAGGAAAAGAACATCTCTCCAAAGTATGCGGATGAAGGGACC 4513
QY 571 alAlaLysLysLysLeuAlaGluAspGlu 580
D 4514 GAGCCGAGGCTGAAGCAAGGGAAGAAAGAA 4542
RESULT 9
US-09-248-796A-183
; Sequence 183, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 183
; LENGTH: 4947
; TYPE: DNA
; ORGANISM: Candida albicans
```

US-09-248-796A-183

Alignment Scores:

```
Pred. No.: 7,37e-06 Length: 4947
Score: 147.50 Matches: 133
Percent Similarity: 36.72% Conservative: 102
Best Local Similarity: 20.78% Mismatches: 243
Query Match: 4.59% Indels: 163
DB: 28
```

SEQ3_K705_COPY_650_1270 (1-621) x US-09-248-796A-183 (1-4947)

```
QY 60 LysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSer 79
D 2698 AAAAATTTGGAACGATTGAATAAATCAATTGAAGAAGAAATTTGTTGGCAACACATC 2757
QY 80 TyrLeuPhe-----AspLysLeuAspHisPheHisSerSer 91
D 2758 TTTTATTTTATTCAAACCAATCCCATTTATTTGACAAAATTTG-----TTTATTTCAAT 2811
QY 92 GluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHisGlu 111
D 2812 CCATATACTAAAAATCAGACAAAATCTGCTCAAGATTGTTCCAGTCTGTGATTCAATTA 2871
QY 112 Phe-----SerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLys 127
D 2872 TTCCCAAGTAAGAGATTTCATCAATTTACTTATCTTCGTGAAGAGTACTTTTTGGTCAA 2931
QY 128 LeuCysLeu---LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThr 146
D 2932 CTTATGATCAGTTGATGCAAAATGATACAGCTAACTCAACAAATTTGGGTGACATCACC 2991
QY 146 ----- 146
D 2992 AAGTTCATTTCACAAACTGGATAGATTTCCTTACTAATTTCAATAATCATATCTCCAG 3051
QY 147 -----LeuPheGlyGluAspHisIleLysLeuSerAsp 157
D 3052 AGACAAACACTTGAAGACACTTTTGGCAAGTTTGTATTACGTATTGTTGATAATGAACAA 3111
QY 158 -----GluGluSerProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnPro 174
D 3112 GTGGATTGTTGAATCAGATCCAAATCAGATCTCAATCAATCAATCAAT-----GATCATGAG 3165
QY 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
D 3166 ATGAAAGTTTATCGTCTGAGTGAAGTCTCGTGATATTATCCACGAGCAGCTATTTCAG 3225
QY 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysLysLysAla 214
D 3226 TTGCCAGAAAGTTAGTAATAAGTTT-----GTGGGCAATTTGATGTCAATTAAGGGAGACT 3279
QY 215 SerAspAspValThrAspProArgValThrAspProValAspAspGluArgLys 234
D 3280 TGCAGTGATTTCCTTCAATGTTA-----AspHisMet-----CAAAAAAT 3312
QY 235 AlaSerGlyLys-----AspHisMet----- 241
D 3313 GTTCTTGGAAATAAAGCCTTGTTCGAAATCCCTGACCATGTGAAATTTGATTGTCGTCAA 3372
QY 242 GlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThr 261
D 3373 GGTATTATTATGTCTCAAGAAAGTTCCCGCAAAATCTGATCAACAAATTTAGCCGTT 3432
QY 262 AspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetCysHisIleSer 281
D 3433 GCTGGGGTTATA-----TTTGTGAACACTACTCTTGGATCT---ATTGTTG 3474
QY 282 GlyIleProGluAsp-----MetLeuAlaSerGlnAspTrpGlyLysIleProAsp 298
D 3475 CAAGTACCAGAGAACTATGGAATTTTGACTGTTAACAATGATACCAGAAAGCTAAAGC 3534
QY 299 ---GluSerGlnArgGluHisThrVal----- 307
```


QY 219 -----ThrAspProArgValThrAspProValAsp-----AspAspGluArg 233
Db 2946 TTGGCCAGAGGTATCAATTTGAACGCTCTCCACAGTCTCTGCAGATGTGGAAAGCCAAAGT 3005
QY 234 LysAlaSerGlyLysAspHis-----MetGlyAlaLeu-----Glu 245
Db 3006 TCTGCTCAAAATTAAGAAAGATACGGCTCTTAGAAGGTTAGCTCTTCTTCAAGAAAGAA 3065
QY 246 SerProLysValIleThr-----LeuGlnSer 254
Db 3066 AGAGACAACTCATTACAGAAATGGACAGGCTCTTTATTGGAAATCAGAGTCTCAGCAGC 3125
QY 255 SerCysLysSer-----SerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGly 272
Db 3126 TCCTGTGAAGTCTAAACTAGCTCTAGAGGGTCTTACTGAAGACAGGAA----- 3176
QY 273 LeuTySerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSer----- 290
Db 3177 -----AAGTTAGTGAAGAAATTTGAATCTTTGAATCTTCTAAGATTGCAGAAAGTACT 3230
QY 291 -----GlnAspTyr 293
Db 3231 GAGTGGCAAGAAACACAGAGGAGCTACAAAAGAGTATGAAATCTTCTGCGAGTCTCTAT 3290
QY 294 GlyLysIleProAspGluSerGlnArgLeuHisThrVal----- 307
Db 3291 GAGAAATGTTAGTAATGAAGCAGAAAGGATTCAGCATGTGTGGTGAAGCTGTGAGGCAAGAG 3350
QY 307 ----- 307
Db 3351 AAACAAGAACTGTATGCGCAAGTTAAGAACGACAGAGGCAAAACAGAGGAGAGACAGAAAG 3410
QY 308 -----LeuLysProLysMetAlaLysLeuCys 316
Db 3411 CAGTTGCAGGAAGCTGACCAAGAAATCGAGGAATGAAGAAAGATGAGAAAGTTGCT 3470
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
Db 3471 AAA-----TCTAAACAGCAGAGAAATCTCTAGAGCTGGAA 3503
QY 337 IleGluAsnHisArgIleTyrGluGlu-----ProAlaThrThrPheGlnAlaPheGln 354
Db 3504 GAAGAGAATGACCGGCTTAGCGCAGAGGTGCACCTTCGACGAGAT----- 3548
QY 355 IleAlaLeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSer 374
Db 3549 -----ACAGCTAAGAGTGTATG-----GAAACA 3572
QY 375 LeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyr----- 393
Db 3573 CTTCTTTCTTCCCAATGCCAGCATGAAGGAAGAACTTGAAGGGTCAAAATGGAGTATGAA 3632
QY 394 -----IleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHis 408
Db 3633 ACCCTTTCTAAGAGTTTTCAGTCTTTAATGCTGTGAGAAAGACTCTCTA-----AGTGAAGAG 3689
QY 409 ThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThr 428
Db 3690 GTTCAAGATTAAAGCATCAGATAGAGGTAAATGTTATCTTAACAGCT----- 3737
QY 429 LysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn----- 446
Db 3738 -----AACCTAGAGGCCACCGAGAAACATGATAACCAACGAATGTCACT 3782
QY 447 -----ThrLysSerMetArgAsnSerSerGluAspGluCysMet----- 460
Db 3783 GAAGAGGGAACACAGTCTATACCAGGTGACACTGAAGAGCAAGCACTCTCTGAGTATGAGC 3842
QY 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
Db 3843 ACAAGACCTACATGTTCCAGATCGGTCCATCAGGAAAGAGTCCCAACCCCTGCTGTAAGT 3902
QY 481 GlyIleLysLysLysTyrLysLys-----GlnValGlnLysLeuValGlnGluHisGlu 498

Db 3903 -----AAGGATTTAGCTCAGCTCAGATGATGAATTAATAACTACCTACAGCAGATTGAT 3953
QY 499 GluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuGluThrSer 518
Db 3954 CAGCTCAAGAAAGAAATTTGCTGATTTAGAGGAGGAGAGCAGAAACAAAGGAATTTAGC 4013
QY 519 LysSerValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
Db 4014 CAGACTTTGAAATTCAGAAA-----AATACCTTACTGAGTCAGATATCAACAAAGGATGCT 4070
QY 539 AspLeuLysLeuAspHisAsnTyrGluArgLysPheAspGluLysSerGluLys 558
Db 4071 GAACTAAATGCTTTCAGGAGGAAGTA-----ACCAAAATGAACCTGTTAAATCAGCAATC 4127
QY 559 AsnGluCysLeuLysSerLeuGluMetHisGluValAlaLysLysLeuAlaGlu 578
Db 4128 CAAGAAGAAGCTTCCAGAGTTACCAACTAAGAGGACAGCAGAGAGAGAAAGATGAT 4187
QY 579 AspGluAlaCysTrpIleAsnArgIle 587
Db 4188 TTGGAAGAGAGGCTTATGAATCAATTA 4214
RESULT 11
US-09-949-016-17304
; Sequence 17304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17304
; LENGTH: 38575
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(38575)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17304
Alignment Scores:
Pred. No.: 0.000306 Length: 38575
Score: 147.00 Matches: 134
Percent Similarity: 36.43% Conservative: 117
Best Local Similarity: 19.45% Mismatches: 250
Query Match: 4.57% Indels: 188
DB: 4 Gaps: 33
SEQ3_K705_COPY_650_1270 (1-621) x US-09-949-016-17304 (1-38575)
QY 18 ValLysHisValGluLysIleLysIleGlu-----SerCysSer 30
Db 4355 CTGAAGAAATAGAGGCTCAGAAAGTAGAGTTAGAAATGAAAGTTAGTTCTACAAAGT 4414
QY 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 4415 GAGCTTACTAAA-----AAATCAGAGAGGATTTTCAG 4447
QY 51 LeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 4448 TTACAGAGCAGATTAATAAACAGGGGTTTGAAGATCGAGAGTCTA-----AAGACAGATATCC 4504


```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12378
; LENGTH: 119153
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(119153)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-12378
```

```
Alignment Scores:
Pred. No.: 0.0022 Length: 119153
Score: 147.00 Matches: 134
Percent Similarity: 36.43% Conservative: 117
Best Local Similarity: 19.45% Mismatches: 250
Query Match: 4.57% Indels: 188
DB: 4 Gaps: 33
```

SEQ3_K705_COPY_650_1270 (1-621) x US-09-949-016-12378 (1-119153)

```
QY 18 VallyHisValGluLysIleLeuGlu-----SerCysSer 30
   |||:|||||
Db 36933 CTGAAGAATAGAGCTGGAAGAAGTAGATTGAATGAACTTAGTTCTACACAGT 36992
QY 31 GluArgThrIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
   |||:|||||
Db 36993 GAGCTTACTAAA-----AAATCAGAGAGGTTATTCAG 37025
QY 51 LeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
   |||:|||||
Db 37026 TTACAAGAGCAGATAAATAACAGGGTTTGAATAATCGAGAGTCTA---AAGACAGTATCC 37082
QY 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer 90
   |||:|||||
Db 37083 CATGAAGCTGAAGTCCATGCCGAAGC---CTCAGCAGAAATGGAA-----AGC 37130
QY 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
   |||:|||||
Db 37131 AGCCAACACTACAATTCGTGSCCTAGAACATCTAAGAGAATTCGAACCT---AAACTGGAT 37187
QY 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLysLeuCysLeu 130
   |||:|||||
Db 37188 GAACCTGCACAAACTCATAGCAAAAG-----GAAGAAGACGCTTAGCTACCTTTCTTGG 37241
QY 131 LeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu 150
   |||:|||||
Db 37242 CAACCTTAGTGAGAAAGAGAGAGCTCTCTCAATAAATACAGACAGAGATAATAGAACAGAA 37301
QY 151 AspHisIleLysLeu-----SerAspGluGluSerProAsn 162
   |||:|||||
Db 37302 GATTTAATTAAGCTCTGCATACACAGCTAGATAATGCAGCCAAAGAGCATGATGAGAGG 37361
QY 163 IlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr 182
   |||:|||||
Db 37362 ATA-----AAGCAGCTTACAGGTGGAACCTTTGTGAATAAGCAAAACCCAGAGAGATT 37415
QY 183 ProGlnArgAsnArgLysArgValGlnTyr-----PheGluGlySerGluAlaSer 199
   |||:|||||
Db 37416 GGAGAGAAAGTAGTAGCAAGCAACAAATAACAAGAAACTGCAAGCTGCCCTTATTTC. 37475
QY 200 ProLysThrGlyAspGlyLysAlaLysArgLysLysAlaSerAspVal--- 218
   |||:|||||
Db 37476 CGAAGAAGCA-----CTAAAGAAACACAAAGTCTCCAGAGGAATTTGTCT 37523
QY 219 -----ThrAspProArgValThrAspProProValAsp---AspAspGluArg 233
   |||:|||||
Db 37524 TTGCCAGAGGTACCAATTGAACGTCTCCCAAGTCTCTGCAGATGTCGAAAGCCCAAGTT 37583
QY 234 LysAlaSerGlyLysAspHis-----MetGlyAlaLeu-----Glu 245
   |||:|||||
Db 37584 TCTGCTCAAAATAAAGAAAAGATACGGTCTTAGGAAGGTTAGTCTCTTCTTCAAGAAGAA 37643
```

```
QY 246 SerProLysValIleThr-----LeuGlnSer 254
   |||:|||||
Db 37644 AGAGACAAACTCATACAGAAATGGACAGTCTTTATTGGAAATCAGAGTCTTCAGCAGC 37703
QY 255 SerCysLysSer-----SerGlyThrAspGlyThrLeuAspGlyAenAspAlaPheGly 272
   |||:|||||
Db 37704 TCCTGTGAAGTCTAAACTAGCTCTAGAGGGTCTTACTGAGACAAAGNA----- 37754
QY 273 LeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSer----- 290
   |||:|||||
Db 37755 -----AAGTAGTGAAGCAATTTGAATCTTTGAAATCTTCTAAGATTGCAGAAAGTACT 37808
QY 291 -----GlnAspTrp 293
   |||:|||||
Db 37809 GAGTGGCAAGAGAAACACAAAGGAGCTACAAAAGAGTATGAAATTTCTTCGACAGTCTAT 37868
QY 294 GlyLysIleProAspGluSerGlnArgArgLeuHisThrVal----- 307
   |||:|||||
Db 37869 GAGATGTTAGTAATGACGAGAAAGGATTCAGCATGTGCTGGAGAGCTGTGAGGCAAGAG 37928
QY 307 ----- 307
   |||:|||||
Db 37929 AAACAAGAACTGTATGCGCAAGTTAAGAAGCACAGACGCAACAAAGAAGGAGACAGAAAG 37988
QY 308 -----LeuLysProLysMetAlaLysLeuCys 316
   |||:|||||
Db 37989 CAGTTGCAGAAAGCTGAGCAAGAAATGGAGAAATGAAAGAAAGATGAGAAAGTTTGTCT 38048
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
   |||:|||||
Db 38049 AAA-----TCTAAACACGACAGAAATCTCTAGAGCTGGAA 38081
QY 337 IleGluAsnHisArgIleTyrGluGlu-----ProAlaThrThrPheGlnAlaPheGln 354
   |||:|||||
Db 38082 GAAGAGAATGACCGGCTTAGGCGCAGAGGTGCACCTCTGCAGGAGAT----- 38126
QY 355 IleAlaLeuSerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSer 374
   |||:|||||
Db 38127 -----ACAGCTAAAGAGGTGTATG-----GAACA 38150
QY 375 LeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyr--- 393
   |||:|||||
Db 38151 CTTCTTTCTTCCAAATCCAGCATGAAGAGAACTTGAAGAGGTCAAAATGGAGTATGAA 38210
QY 394 -----IleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHis 408
   |||:|||||
Db 38211 ACCCTTTCTAAGAGTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTA---AGTGAAGAG 38267
QY 409 ThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThr 428
   |||:|||||
Db 38268 GTTCAAGATTTAAGCNCATCAGATAGAAGGTATGTATCTAACAAGCT----- 38315
QY 429 LysLeuValAsnGluSerLeuSerGlyAlaThrValArgAspLysIleAsn----- 446
   |||:|||||
Db 38316 -----AACCTAGAGCCACCGAGAAACATGATACCAACCAAGATGTCACT 38360
QY 447 -----ThrLysSerMetArgAsnSerSerGluAspGluGluCysMet----- 460
   |||:|||||
Db 38361 GAAGAGGGAACACAGTCTATACCGGTGAGACTGAGAGAGAGAGACTCTCTGAGTATGAGC 38420
QY 461 ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
   |||:|||||
Db 38421 ACAAGACCTTACATGTTTCAGATCGGTTCCATCAGCGAAGAGTGCCCAACCTGCTGTAAGT 38480
QY 481 GlyIleLysLysLysTyrLysLys-----GlnValGlnLysLeuValGlnGluHisGlu 498
   |||:|||||
Db 38481 -----AAGGATTTACGCTCACATGATGAATTAATACTACCTACAGAGATGATG 38531
QY 499 GluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysLeuGlnLysLeuGluThrSer 518
   |||:|||||
Db 38532 CAGCTCAAGAAAGAAATTCCTGATTTAGAGGAGAGAGACAGAAACAAAGGAATTTAGC 38591
```

Qy	519	LyseRValGluAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly	538
Db	38592	CAGCTTTAGAAATCAGAAA---AATACCTTACTGAGTCAGATATCAACAAGAGTGT	38648
Qy	539	AspLeuValLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys	558
Db	38649	GAACTTAAATGCTTCAGAGGAGTA---ACCAAAATGAACTCTTAAATCAGCAATC	38705
Qy	559	AsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGlu	578
Db	38706	CAAGAAGAACTCTCCAGAGTTCACAACTAAAGGAGACAGCAGAGAAGAGCAAGATGAT	38765
Qy	579	AspGluAlaCysTrpIleAsnArgIle	587
Db	38766	TTGGAAGAGAGGCTTATGAATCAATTA	38792
RESULT 13			
US-09-917-254-40			
; Sequence 40, Application US/09917254			
; Patent No. 6703204			
; GENERAL INFORMATION:			
; APPLICANT: Mutter, George			
; APPLICANT: Baak, Jan			
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer			
; FILE REFERENCE: B0801/7224(JRV)			
; CURRENT APPLICATION NUMBER: US/09/917,254			
; CURRENT FILING DATE: 2001-07-27			
; PRIOR APPLICATION NUMBER: US 60/222,093			
; PRIOR FILING DATE: 2000-07-28			
; NUMBER OF SEQ ID NOS: 102			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 40			
; LENGTH: 5574			
; TYPE: DNA			
; ORGANISM: Homo Sapiens			
US-09-917-254-40			
Alignment Scores:			
Pred. No.:	3 08e-05	Length:	5574
Score:	143.00	Matches:	142
Percent Similarity:	35.68%	Conservative:	96
Best Local Similarity:	21.29%	Mismatches:	229
Query Match:	4.45%	Indels:	202
DB:	4	Gaps:	32
SEQ3_K705_COPY_650_1270 (1-621) x US-09-917-254-40 (1-5574)			
Qy	22	GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyr:SerVal	41
Db	2554	CAGAAGCTGCNACTTGAGAAGGTCACGGCTGAGGCCAAGATCAAGAAACTG	2604
Qy	42	CysThrValGluGluLysAlaLysIleLeuAlaArgGlnAsnLysArgGlnAsnLysAla	61
Db	2605	-----GAGATGAGATCTCTGTCATGATGATCAGAACATAAATCTATCAAAA	2652
Qy	62	ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu	81
Db	2652	-----	2652
Qy	82	PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu	101
Db	2653	-----GAA	2655
Qy	102	GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSertyeGlyGly	121
Db	2656	CGAAAACTCTTCAGGAGAGGATTAGTACTTAACGACAAATCTTTGCAGAA	2706
Qy	122	GluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyr	141
Db	2707	GAGGAGAAAGGCCAAGAACTCTTACCAGAGCTGAAAGAACAGCAT	2751
Qy	142	SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerPro	161

Qy	469	erThralaThr-----ArgAspIleGluLysThrI	479
Db	3755	CCACTCTCAACATCCAGCTCTCCGACTCGAAGAAGAGCTGGAGGCTTTTCGCGACACCG	3814
Qy	479	leSerGlyIle-----LysIlystYsTyrlYsLysGlnValGlnLysLeuValGlnG	496
		:::	
Db	3815	TGGAAAGCTCTGGAAGAGGGGAAGAGAGGTTCAGAAAGAGAGATCGAGAACTTCACCCAGC	3874
		:::	
Qy	496	luHisGluGluLysLysMetGluLeuLeuAsnMetTyAlaAspLysLysGlnLysLeuG	516
Db	3875	AGTACGAGGAGAAG-----GCGGCGCTTATGATAAACTGG	3910
Qy	516	luThrSerLysSerValGluAlaValIleArgIleThrCysSerArgThr-SerThrG	536
Db	3911	AAAAGACCAAG-----AACAGGCTTCAGCAGG	3937
Qy	536	lnValGlyAspLeuLysLeu---LeuAspHis-----A	546
		:::	
Db	3938	AGCTGGACGACCTGGTTGTTGATTGTCACAACACGAGCGCAACTCGTGTCCAACCTGGAAA	3997
Qy	546	snTyrgluAArgLysPheAspGlu---IleLysSerGluLysAsnGluCysLeuLysSerL	565
Db	3998	AGAAGCAGAGGAAATTTGATCAGTTGTACCGGAGAGAAAAACATCTCTCCAAAATACG	4057
Qy	565	euGluGlnMetHisGluVal---AlaLysLysLysLeuAlaGluAspGluAla-Cys---	582
		:::	
Db	4058	CGGATGAGAGGGACAGAGCTGAGCGAGAGCCAGGGAGAGGAACCAAGGCGCTGTCCC	4117
Qy	583	TrpIleAsnArgIleLys---SerTrpAlaAlaLysLeuLysValCysValProIleGln	601
Db	4118	TGGCTCGGCGCCTTGAAGAGCGCTTGGAAGCCAAAGAGGAACTCGAGCGGACCAACAAA	4177
Qy	602	SerGlyAsnAsnLys	606
Db	4178	TGCTCAAGACCGAAA	4192

RESULT 14

US-09-949-016-1240

03-03-943-010-1240
: Sequence 1240, Application US/09949016

Sequence 1230, application 00/05533010
; Patent No. 6812339

FACEID NO. 0812333
: GENERAL INFORMATION:

APPLICANT: VENTER,

1. TITLE OF INVENTION:

TITLE OF INVENTION:

; FILE REFERENCE: CL0

CURRENT APPLICATION

CURRENT FILING DATE

PRIOR APPLICATION NUMBER

PRIOR FILING DATE:

1. PRIOR APPLICATION NUMBER

PRIOR FILING DATE:

PRIOR APPLICATION N

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NO

; SOFTWARE: FastSeq f

; SEQ ID NO 1240

; LENGTH: 6861

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-1240

Alignment Scores:

Pred. No.:

Score: _____

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

SEQ3_K705_COPY_650_127

1
2
3
4
5
6
7

Qy 22 GluLysII

Db	2987	CAAGAGCTGCAACTTGGAGAGGTCACGGCTGAGCGCCAAAGCAATCAAGAAATCGT-----	3037
Qy	42	CysThrValGluGluLysAlaLeuIleLeuAlaAArgGlnAsnLysArgGlnAenLysAla 61	
Db	3038	-----GAGGATGAGATCTCGTGGTCATGGATGATCAGACAAATAAATCATCAAAA---	3085
Qy	62	ValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu 81	
Db	3085	-----	3085
Qy	82	PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGlu 101	
Db	3086	-----GAA 3088	
Qy	102	GlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 121	
Db	3089	CGAAAACTCTCTGAGGAGGATTAGTGACTTAACGACAAATCTTCAGAA-----	3139
Qy	122	GluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyr 141	
Db	3140	GAGGAGAAAGGCCAAGAAATCTTACCAAGCTGTAAAAACCAAGCAT-----	3184
Qy	142	SerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerPro 161	
Db	3185	-----CAATCTATGATTTCAGAACTGGGAAGTCGGCTAAAGAGAGAGAGAGAGC 3235	
Qy	162	AsnIlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAsp 181	
Db	3236	CGACAGAGCTGCAGAAAGCTG-----AAACGGAAAGCTGGAGGCTGATGCCAGCGAC 3286	
Qy	182	ThrProGlnArgAsnArg-----LysArgValGlnTyrPheGluGly 195	
Db	3287	TT-CCACGACGATCGTGACCTCCAGCGGCAGATCGCAGAGCTCAAGATCAGCTGGC 3345	
Qy	196	SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysArgLysLysAlaSer 215	
Db	3346	CAAGAGGAGGAGGAGCTGCAGCGCGCCCTGGCCAGCTTGACGATGAATCGCTCAGAA 3405	
Qy	216	AspAspVal-----ThrAspProArgValThrAspProVal-Asp-----	229
Db	3406	GAACAATGCCCTGAAGAAAGATCCGGAGCTGGAGGGCCACATCTCAGACCTCCAGAGGA 3465	
Qy	230	AspAspGluArgLysAlaSerGlyLys-----As 239	
Db	3466	CCTGGACTCAGACGGCGGCCAGGAACAGGCTGAAAGCAGACGAGACCTCGCGCA 3525	
Qy	239	pHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSe 259	
Db	3526	GGAGCTGGAGGCCCTAAAGACAGAGCTGGAAGACACACACTGGACAGCACACGCCAC-TCAGC 3584	
Qy	259	rgLysThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMetGlySerHi 279	
Db	3585	AGAGGCTCAGGGC-----CAAGAGGGAGCAGG 3611	
Qy	279	SileSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIlePro-AspG 299	
Db	3612	AGGTGACGTGCTGAAGAAGGC-----CCTGGATG 3641	
Qy	299	lueSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValL 319	
Db	3642	AAGAGACGGCGTCCCATGAGGCTCAGGTCCAGGAGATGAGGAGAGAAACACGACACGGCGG 3701	
Qy	319	euhHisLeuSerAspAlaCysThrSerMetValGlyAsnPhe-----	332
Db	3702	TG-----GAGGAGCTCACAGACACCTTGACGATTCAGAGGGCCCAAGGGCAACC 3752	
Qy	333	-----LeuGluTyrValIleGluAsnHisArgIleTyrGluGluP 346	
Db	3753	TAGACAAAGAATAAGACAGACGCTGGAGAAA-----GAGAACGCAGACCTGGCGGGAGC 3806	
Qy	346	roAlaThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuValL 366	

Qy 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41


```
Db 3526 GAGCTCGAGCGCCCTAAGACAGAGCTGGAGACACACTGGACAGCACGCCAC-TCAGC 3584
Qy rGlyThrAspGlyThrLeuAspGlyAenAspAlaPheGlyLeuTyrSerMetGlySerHi 279
Db 3585 AGGAGCTCAGCGC-----CAAGAGGGAGCAGG 3611
Qy sIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTyrGlyLysIlePro-AspG 299
Db 3612 AGGTGACGGTGTCTGAAGAGC-----CCTGATG 3641
Qy 299 luSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuLysGlnVal 319
Db 3642 AAGAGACGGCTCCCATGAGGCTCAGGTCAGGATGAGGATGAGGACAAACACGACAGCGG 3701
Qy 319 euHisLeuSerAspAlaCysThrSerMetValGlyAenPhe----- 332
Db 3702 TG-----GAGGAGCTCAGAGCAGCTTGACGATTCAGAGGGCCCAAGCGGAC 3752
Qy 333 -----LeuGluTyrValIleGluAenHisArgIleTyrGluGluP 346
Db 3753 TAGACAAGAATAAGACAGCCTGGAGAAA-----GAGAACGACAGACCTGGCCGGGAGC 3806
Qy 346 rolaThrThrPheGlnAlaPheGlnIleAlaLeuSerTyrIleAlaLeuVal 366
Db 3807 TCGGGTCTCGGCCAGGCC-----A 3827
Qy 366 ysGlnIleLeuSerHisLysGluSerLeuValArgAlaAen---SerGluLeuAlaPheL 385
Db 3828 AGCAGAGGTGGGAACATAGAAGAAGAGCTGGAGCGCGAGTGCAGGAGCTGCAGTCCA 3887
Qy 385 ysCysSer-----ArgValGluVal---AspTyrIleTyrSerIleLeuS 399
Db 3888 AGTGACAGCATGGGAGCGGCCCGCGCGAGCTCAATGACAAAGTCCACAAAGCTGCAGA 3947
Qy 399 erCysMetLysSerLeu-----PheLeuGluHisThrGlnGly-----LeuG 413
Db 3948 ATGAAGTTGAGACGCTCACAGGATGCTTAACAGGCGCGAGGGGAAGGCCATTAAAGCTGG 4007
Qy 413 lnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnG 433
Db 4008 CCNAGAGCTGGCGGTCCCTCAGTCTCCAGCTCCAGGACCCAGGAGCTGCTTCAAGAAG 4067
Qy 433 luSerLeuSerGlyAlaThrValArgAspGluLysIleAenThrLysSerMetArgAsnS 453
Db 4068 AAACCCGGCAGAGCTCAACGTCTTACGAAGCTGCGCCAGCTGGAGGAGGCGGGAACA 4127
Qy 453 erSerGluAsp-----GluGluCysMetThrGluLysArgCys-----SerHisTyrS 469
Db 4128 GCTGCAAGACCAAGCTGGAGCAGGAGATGGAGGCCAAGCAGAACCTGGAGCGCCACATCT 4187
Qy 469 erThrAlaThr-----ArgAspIleGluLysThrI 479
Db 4188 CCACTCTCAACATCCAGCTCTCCGACTCGAAGAAAGAGTGCAGGAGCTTTCAGGACCCG 4247
Qy 479 leSerGlyIle-----LysLysLysTyrLysLysGlnValGlnLysLeuValGlnG 496
Db 4248 TGGAAAGCTCTGGAAGAGGGAAGAGAGGTTCCAGAAGGAGATCGAGAACCTCACCCAGC 4307
Qy 496 luHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGlnLysLeuG 516
Db 4308 AGTACGAGGAGAG-----GCGGCGCTTATGATATAACTGG 4343
Qy 516 luThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrC 536
Db 4344 AAAGACCAAG-----AACAGCTTTCAGCAGG 4370
Qy 536 lnValGlyAspLeuLysLeu---LeuAspHis-----A 546
Db 4371 AGCTGGACGACCTGGTTGTTGATTGGACAACACCGCGCAACTCGTGTCCAACCTCGAAA 4430
Qy 546 snTyrGluArgLysPheAspGlu---IleLysSerGluLysAenGluCysLeuLysSerL 565
```

```
Db 4431 AGAAGCAGAGGAATTTGATCAGTTGATCCCGAGGAGAAAAAATCTCTTCCAAATACG 4490
Qy 565 euGluGlnMetHisGluVal---AlaLysLysLysLeuAlaGluAspGluAla-Cys--- 582
Db 4491 CGGATCAGAGGGACAGAGCTGAGCGCAGAACCCAGGAGAGGAACCAAGCCCTGTCCC 4550
Qy 583 TrpIleAsnArgIleLys---SerTyrAlaAlaLysLeuLysValCysValProIleGln 601
Db 4551 TGGCTCGGGCCCTTTGAAGAGGCTTGGAAAGCCCAAGAGGAACCTCGAGCGGACCAACAAA 4610
Qy 602 SerGlyAsnAsnLys 606
Db 4611 TGCTCAAGCCGAAA 4625
```

Search completed: July 12, 2005, 00:10:17
Job time : 406.333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 12:27:00 ; Search time 983.333 Seconds
(without alignments)
3964.922 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNKHFGSSNISQNPADV 621

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6330945_seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2_1/USPTO_spool/ROBINSON929/runat_11072005_102153_29226/app.query.fasta_1.2325
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -LOCAL=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000
-USER=ROBINSON929 @cgn 1.1370 @runat_11072005_102153_29226 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	591.5	18.4	6409	19	US-10-437-963-76223 Sequence 76223, A
2	491.5	15.3	8425	19	US-10-437-963-69891 Sequence 69891, A
3	456.5	14.2	5865	19	US-10-437-963-76215 Sequence 76215, A
4	336.5	10.5	1197	20	US-10-425-115-159032 Sequence 159032, A
5	192.5	6.0	487	18	US-10-424-599-12966 Sequence 12966, A
6	166	5.2	7764	10	US-09-814-353-21548 Sequence 21548, A
7	166	5.2	7764	17	US-10-341-434-60 Sequence 60, Appl
8	166	5.2	7764	21	US-10-651-237-27 Sequence 27, Appl
9	166	5.2	7764	21	US-10-782-413-27 Sequence 27, Appl
10	164	5.1	3045	17	US-10-369-493-26897 Sequence 26897, A
11	161	5.0	661	20	US-10-425-115-145108 Sequence 145108, A
12	159	4.9	5119	21	US-10-956-157-4967 Sequence 4967, App
13	159	4.9	5857	10	US-09-873-367C-305 Sequence 305, App
14	159	4.9	5857	21	US-10-843-641A-305 Sequence 305, App
15	159	4.9	5857	21	US-10-831-704-79 Sequence 79, Appl
16	159	4.9	5857	21	US-10-956-157-975 Sequence 975, App
17	159	4.9	5857	21	US-10-956-157-976 Sequence 976, App
18	157.5	4.9	5883	21	US-10-956-157-859 Sequence 859, App
19	157.5	4.9	7274	21	US-10-696-909A-48 Sequence 48, Appl
20	157.5	4.9	7396	17	US-10-028-248A-35 Sequence 35, Appl
21	157.5	4.9	7396	17	US-10-107-782-35 Sequence 35, Appl
22	156.5	4.9	6354	15	US-10-084-817-158 Sequence 158, App
23	156	4.9	6364	13	US-10-071-766-140 Sequence 140, App
24	156	4.9	7659	19	US-10-437-963-27251 Sequence 27251, A
25	154.5	4.8	6965	20	US-10-335-053-229 Sequence 229, App
26	154.5	4.8	7694	16	US-10-096-534-34 Sequence 34, Appl
27	154.5	4.8	7695	11	US-09-968-007A-462 Sequence 462, App
28	154.5	4.8	7695	21	US-10-843-641A-6932 Sequence 6932, Ap
29	154.5	4.8	8063	10	US-09-814-353-21776 Sequence 21776, A
30	153.5	4.8	4551	18	US-10-675-072A-42 Sequence 42, Appl
31	150	4.7	5847	16	US-10-032-585-6611 Sequence 6611, Ap
32	148	4.6	5641	16	US-10-032-585-6646 Sequence 6646, Ap
33	147.5	4.6	2434	20	US-10-425-115-109076 Sequence 109076
34	147.5	4.6	4944	16	US-10-032-585-6572 Sequence 6572, Ap
35	147.5	4.6	5676	19	US-10-437-963-58717 Sequence 58717, A
36	143	4.4	3388	9	US-09-954-456-1602 Sequence 1602, Ap
37	143	4.4	3388	9	US-09-967-768A-245 Sequence 245, App
38	143	4.4	3388	9	US-09-954-531-988 Sequence 988, App
39	143	4.4	3388	10	US-09-954-531-1382 Sequence 1382, Ap
40	143	4.4	3388	10	US-09-873-367C-85 Sequence 85, Appl
41	143	4.4	3388	18	US-10-240-425-1200 Sequence 1200, Ap
42	143	4.4	3388	21	US-10-843-641A-85 Sequence 85, Appl
43	143	4.4	3388	21	US-10-843-641A-2055 Sequence 2055, Ap
44	143	4.4	3388	21	US-10-843-641A-2449 Sequence 2449, Ap
45	143	4.4	3388	21	US-10-843-641A-4629 Sequence 4629, Ap

ALIGNMENTS

RESULT 1
US-10-437-963-76223
; Sequence 76223, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76223
; LENGTH: 6409
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76236C.1
US-10-437-963-76223

Alignment Scores:
Pred. No.: 1,71e-50 Length: 6409
Score: 591.50 Matches: 175
Percent Similarity: 45.21% Conservative: 127
Best Local Similarity: 26.20% Mismatches: 239
Query Match: 18.39% Indels: 127
DB: 19 Gaps: 20

SEQ3_K705_COPY_650_1270 (1-621) x US-10-437-963-76223 (1-6409)

QY 3 AspAlaPheIleuPheGlySerLeuAsnProSerHisAspValLysHisValGlu 22
DB 2263 GATGCCATCATATATTTGTTAGTGGATTCGAAATCCAACCAATGACTTGGAGTTCTCCAG 2322

QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
DB 2323 AGGATCAGTATAGAGTCTCAATCTGAATGTGTCCTATTTTCGCTGTGTTATTCGTTCTGT 2382

QY 43 ThrValGluGluLysAlaLeuLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
DB 2383 ACGGTGGAGGAAAGACTCTTATCTACTAGCGAAGCATGATCATATCTTTCGACGCAACGTC 2442

QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGluValSerTyrLeuPhe 82
DB 2443 CAGATGTGTAATCGCTATCGTGAGCCATTCCTGCTTAGTTGGGTGTCATCGTTCTCTTC 2502

QY 83 AspLysLeuAspHisPheHis-----SerSerGluThrProAspSerGly 97
DB 2503 AATAGACTTGGAGGTTCCAAAAGCATGACTACTCCAGTAAAGATTCTCGAGGATGATGT 2562

QY 98 ValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSer 117
DB 2563 CTC-----TTCATGATAATGTATTTTGGAGTTTTCGAGTTTCGACGAAACTATCC 2607

QY 118 SerLysGlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAla 137
DB 2608 ACCAAGTTGAAGCCGCACTAAATGGAAATGCAATGTCAGTCAATATCTCGAGCTCAGCAGAT 2667

QY 138 GlnGlyThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSer 156
DB 2668 GGATCGTTTATTCTAGAGATATGCTGTTATAAGCGAGAGGAGGGAATATCTGCAGTT 2727

QY 157 AspGluGluSerProAsnIle-----PheTrpSerLysLeuLeuGluClyLysAsnPro 174
DB 2728 GATGGTGAATTCGCAAGTTCTCGACCTTCGTCATAATTTTACTAGTGGGAGGTCCTCT 2787

QY 175 MetTrpLysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu 194
DB 2788 CATTCGCAATATATATCTGAGCCAGTGCAGAGAACCGTAGGAGATACAAAATATGGAA 2847

QY 195 GlySerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArgLysAla 214
DB 2848 GACCAATGAGAATACCTTCAGAGAAACTGTAAGCAATAATGAACGTAGAAATATC 2907

QY 215 SerAspAspValThrAspProArgValThrAspProValAspAspAspGluArgLys 234
DB 2908 GGAGAA-----ATCATGGAT----- 2922

QY 235 AlaSerGlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSer 254
DB 2923 -----TCATCTCCAGAGATCTCTCTGTCATGCAATTTTGAAT 3891

```

```

QY 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
DB 2953 -----AATGATGCTGTG----- 2964

QY 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAsp----- 292
DB 2965 -----CTGCTGAAATAGTACAGCATCCAGTTCTCATGAA 3000

QY 293 -----TrpGlyLysIleProAspGluSer-----GlnArg 302
DB 3001 ACATCAGTTGATGACACTTGGCAAGCTAGGGCGAGAAAGCCTTCAGGGCACACAGAA 3060

QY 303 ArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSer 322
DB 3061 GGTCTTTCACCCAACTCAAGCCAGAGCTGTCAAAAGTTGTATGAATTTGTGAATTTCCG 3120

QY 323 AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIle 342
DB 3121 GAGACTGTCAATGTTTATGTGAGGAATTACTTGAATTACATTTTGAAGTAATCATCAAGTC 3180

QY 343 TyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAla 362
DB 3181 AGTCAGAGCCAAAGGCGATATTGTCATGCAATTCACATTCCTGTTGGCGTGTCTGT 3240

QY 363 LeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeu 382
DB 3241 TCTCTCTTAAAGCATAAAGATCAATCGAAGAGAGTCACTTGCCTTCTGTGAGAAACTTG 3300

QY 383 AlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLys 402
DB 3301 AACTATGATGCGATGAAGTGTCTGCTGAATATGTTTATGAGAAAGTAAAGATTCCTCAAG 3360

QY 403 SerLeuPheLeuGluHisThr-----GlnGlyLeuGlnPheAspCysPhe 417
DB 3361 AAAAAAGTTTTCGCTAGAGCAAGTGAACACAGCAGAGCAATCTACTCCAGTAAT 3420

QY 418 GlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSer-----Leu 435
DB 3421 AACACATCATCTTACAAACACAGACCTCACCGAAATTTGAGAAAGTGCAGGATCAATTTGT 3480

QY 436 SerGlyAlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGlu 455
DB 3481 CATCAGGTAAACCAATGATGTTGTTGAAATGTTTTCACATGAGAGAGCTCCACAT 3540

QY 456 Asp-----GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThr 472
DB 3541 GATATCTTTCAGCGAGAGATGATATTAGAACAGAG----- 3576

QY 473 ArgAspIleGluLysThrIleSerGlyLysLysLysLysLysLysGlnVal----- 490
DB 3577 -----GAATTGATATCTGTTCTAGAACTCATAGAGAAACATGTTTAAAG 3624

QY 491 GlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAsp 510
DB 3625 GATGAATCTTCTGAAAGAAATCACAGAAAGAAATTAATTTAATTAACATGTTTCTCC 3684

QY 511 LysLysGlnLys-----LeuGluThrSer 518
DB 3685 TTAAGAGAAAGAAATATTTCAGGATAAACCAAGAAACGAGACCACACTGTTAGACATGCAC 3744

QY 519 LysSerValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGly 538
DB 3745 AAGCAAAAGGAGTAGTGCAGAAAGCTGCAGAAACATGCAAT----- 3783

QY 539 AspLeuLysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLys 558
DB 3784 -----TTGCTGCGGAACATCTTCGTAAGGTCACTTGAT-----TCAGAGGACAGG 3831

QY 559 AsnGluCysLeuSerLeuGlnMetHisGluValAlaLysLysLysLeuAlaGlu 578
DB 3832 GATGCCACGGTGAAGTAATTAATGATGTTTACTTTGCTTCTGTATGCAATTTTGAAT 3891

```

QY 579 AspGluAlaCysTrpIleAsnArgIleLys-----SerTrpAlaAlaLys 593
 Db 3892 CACATGAGATGCCAGCATATAACAAGTTGAAGATGCAGCAATCAACTTCATGGAACAAGGAG 3951
 QY 594 LeuLysValCys---ValProIleGlnSerGlyAsnAsnLysHisPheSerGlySer 612
 Db 3952 TTACAATTGAGGAATTTCTTCAGCAAGCAAAATCTGGCCACTTAGATCGTAGCTTT 4011
 QY 613 AsnIleSerGlnAsnAlaProAsp 620
 Db 4012 GATCAACAATAATCTCTTACCAGAT 4035

RESULT 2

US-10-437-963-69891/c
 ; Sequence 69891, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 69891
 ; LENGTH: 8425
 ; TYPE: DNA
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_70516C.1
 US-10-437-963-69891

Alignment Scores:

Pred. No.: 8,19e-40 Length: 8425
 Score: 491.50 Matches: 176
 Percent Similarity: 40.54% Conservative: 107
 Best Local Similarity: 25.21% Mismatches: 256
 Query Match: 15.28% Indels: 159
 DB: 19 Gaps: 20

SEQ3_K705_COPY_650_1270 (1-621) x US-10-437-963-69891 (1-8425)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
 Db 5785 GATGCCATCATATATATGTTCTGCAACAACCCACTAAACGACTTGAAGGCTCTTCAA 5726
 QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
 Db 5725 AAGATCAAGATAGAGTCACAAATTTGACGCTGTGAGCAATTTTCGCTGTGACACTCTCTTT 5666
 QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
 Db 5665 ACAGTTGAGGAGAGAGCTGTGTCTTCGAGGAGGCGCAATGTTATTGATAACAACATA 5606
 QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
 Db 5605 CAAGATCTAAGGACTAGTTTAAACATCTCTTACTCAGATGGGTGCTGCATCTCTTCTC 5546
 QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
 Db 5545 AGTAGACTTGATGGTCCAGCAGATGACCATGCAAGTAAAGTTCTGAAATGGAGAGA 5486
 QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlu 122
 Db 5485 CACTTTATTGATGAAGTATCGTTGAGTTCTTAAACAAGTTATCCACAACACTGTTGAGAC 5426

QY 123 GluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
 Db 5425 AGCACTGAAGTACACAGGAAGTCCATAAGCAAGCTAATATAGTGGGGAACATATATCA 5366
 QY 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsn 162
 Db 5365 AGAATAATTAATCTGATGGTGAAAGAGGGAATCTCTGCTGTGAGGATATATCCAGCT 5306
 QY 163 IlePheTrpSerLysLeuLeuGlyLysAsnProMetTrpLysTyrProSerAspThr 182
 Db 5305 GAGTTCCTGGTTAAATTTATTGGATGAAGATCCCTCATGCTGAGCTGATATCTGAGCCA 5246
 QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGlnLysSerProLysThr 202
 Db 5245 CTGCAATCAAGATTACAAAGTCCAGACCATGAT-----GAGGTCAATGCTCCGGCT 5192
 QY 203 GlyAspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspVal-----Thr 219
 Db 5191 GAAGAAATCAATGAAGCCAGAAAGACATAGAAAGTTGGTGAGATATGGGTTCACTCT 5132
 QY 220 AspProArgValThrAspProProValAspAsp-----AspGluArgLysAlaSer 236
 Db 5131 TCAAAAGTTGTAATCTGACAAAAGTAACGATGATGCTTGCCTGATATTTGCTACTACATCT 5072
 QY 237 GlyLysAspHisMetGlyAlaLeuLeuSerProLysValIleThrLeuGlnSerSerCys 256
 Db 5071 GGCCTC-----GCTCTTCAA---CCAGTTGATGTCACGACGCAAAATCAAGTT 5027
 QY 257 LysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyrSerMet 276
 Db 5026 CAATCTGAAGGTGGGAAGCCCTGATG-----5000
 QY 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGlyLysIle 296
 Db 5000 -----5000
 QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
 Db 4999 -----AGTACACCCCAAGAAATCTTTCATGCCCAAAATGAAGCAGAGTTGTCGAAATTAATT 4946
 QY 317 GlnValLeuHisLeuSer-----322
 Db 4945 AAGGTGTACAACTACCGAATACATACAGTTTAATGGAGCGAAATATAGACAGCAAGTT 4886
 QY 323 -----AspAlaCysThrSerMetValGlyAsnPheLeuGluTyrValIleGlu 338
 Db 4885 CTCATTGAAGTAGACAATGTTACGTTGCTGGTTGAGCAGTTCTTTGAAATATCTTTGAAT 4826
 QY 339 AsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSer 358
 Db 4825 AATCATGTAGTTGTTTCAGGAGCCAAAGTACATATTTTCATCGTTGAACATAGCTTTGTTGT 4766
 QY 359 TrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArgAla 378
 Db 4765 TGCGAGTTGCCCTCTATTCATAACTTTAAGTGGATCATAAAGAGTCACCTTGCCCTTGCT 4706
 QY 379 AsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLeu 398
 Db 4705 GAAAAAGGTTAAAAATACGAGTGCATAGGAGCTGCGAAGGTTGGTCTATGACAGCCTA 4646
 QY 399 SerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPheGly 418
 Db 4645 -----AAGAGAAAAATTTCCAAAGAAAGCAGGTCACACTGGCAGCAATTCGTAATCA 4595
 QY 419 ThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGluSerLeuSerGlyAla 438
 Db 4594 ACA-----TCAGTAGAAAAAGACCAACCTTTCACAGCAGGAGACCTCTAAT--- 4550
 QY 439 ThrValArgAspGluLysIleAsnThrLys-----448
 Db 4549 ATCTTGAGAAATGACCACATTTTCCCTAAGCAGAGATGCGATCTTTCATGATATTTTCATG 4490
 QY 449 -----SerMetArgAsnSerSerGluAspGluGlu 458

```
Db 4489 AATGTCACCTTCAAGAGGCTCATTTGTCGCACAAATGTTTCAGAGCAGGAA 4430
Qy 459 CysMet-----ThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 4429 CTGATAGCTGTTCCAGGAATCACCATGGATGC---CATTTTCACTGATCACTTCT 4373
Qy 475 -----IleGluLysThrIleSerGlyLysLysLysLysLysLysGlnValGlnLys 492
Db 4372 GACATTGTGGAGAAAAGAAATCAACTTAATTGCAATGTTTTCTCTGTAAGAGAAATACCGT 4313
Qy 493 LeuValGlnGluHisGluLysLysMetGluLeuLeuLeuAsnMetTyrAlaAspLysLys 512
Db 4312 ATCTTTGTAATAACAACAGTCTCAGATTTTCAGATTTTACAGAAATACACACAAAATAAACT 4253
Qy 513 GlnLysLeuGluThrSerLysSerValGluAlaAlaValIleArgIleThrCysSerArg 532
Db 4252 GCCAGCTGGAACA-----GTGTGCAAT--- 4229
Qy 533 ThrSerThrGlnValGlyAspLeuLysLysLeuLeuAspHisAsnTyrGluArgLysPheAsp 552
Db 4228 -----TTAGTTCTGGAACATATTTGTAGAAAGTCATGCTGAT 4193
Qy 553 GluIleLysSerGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisGluValAla 572
Db 4192 -----GTAGAGCCCGAAATGACACAAATAAAACAAACTGTTTCAGTGGTTCACTATGCTG 4139
Qy 573 LysLysLysLeuAlaGluAspGluAlaCysTyrIleAsnArgIleLysSer----- 589
Db 4138 ATGTATGCAATTTCTGGACACATGAGATCCAGCACAGCAAACTTGATGCTGCACTGCAATCT 4079
Qy 590 -----TrpAla-----AlaLysLeuLysValCysValProIleGlnSerGly 603
Db 4078 AATACATGGGCTGAAGAGCGACAGCTGAAGGAAAAACTTTGTTGGAAGCAAAATCAGGC 4019
Qy 603 ----- 603
Db 4018 CAATTAGATCACACCTTTTGATCAGCAGATTTGCTTACCAGATTCAAACTTTGTTATGCNA 3959
Qy 604 -----AsnAsnLysHisPheSerGlySerSer 612
Db 3958 GAATTCATCCATTTGAGGAACAAGTAGCAATTCCTCATGTCAGTGGAGCGCT 3905

RESULT 3
US-10-437-963-76215
; Sequence 76215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76215
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76229C.1
US-10-437-963-76215
```

```
Percent Similarity: 41.74% Conservative: 118
Best Local Similarity: 24.02% Mismatches: 229
Query Match: 14.19% Indels: 159
DB: 19 Gaps: 24

SEQ3_K705_COPY_650_1270 (1-621) x US-10-437-963-76215 (1-5865)

Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db 2263 GATGCCATCATCATATATTGTTAGTATTGGAATCCAAACCAATGACTTGAGAGTTCTCCAG 2322
Qy 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 2323 AGGATCAGTATAGACTCTCAATCTCAATGTGTGCCTATTTTCGCTTGTGCTTGT 2382
Qy 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
Db 2383 ACGTGGAGGAAAGACTCTTATCTG----- 2409
Qy 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeu--- 81
Db 2410 -----CGAAGCATG-----ATCATATTCTTGACACAGCAACGTCAGAAATACTT 2451
Qy 82 -----PheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSer 99
Db 2452 GAGGAGTTCCAAAAGCATGAC---TACTCCAGTAAAGATTCTGAGGATGATGGTCTC--- 2505
Qy 100 PheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLys 119
Db 2506 -----TTCATGAATAATGTAATTTTGGAGTTTGGCAGCGAAACTATCCACCAAC 2553
Qy 120 GlyGlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGly 139
Db 2554 GTTGAAGCCAGCACCTAAAATGGAATGCAATGCACTCATATCTCGAGCTCAGCAGAGTGATCG 2613
Qy 140 ThrTyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGlu 158
Db 2614 TTTTATTCTAGAGATATTGCTGTTATAAGCGAGAGGGGAATATCTCGACTTGTATGCT 2673
Qy 159 GluSerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrp 176
Db 2674 GATTTGCCAAAGTTCTGGACCTTCTGGTCAAAATTTACTAGTGGGAGGTCTCTCATTTGG 2733
Qy 177 LysTyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySer 196
Db 2734 CAATATATATCTGAGCCAGTGCAGAGGAAACCGTAGAAGATACAAAATATGGAAGACCAA 2793
Qy 197 GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysArgLysLysAlaSerAsp 216
Db 2794 ATGAGATACCTGCAGAGAAACTGATGAAGCAATATGAACACGTAGAAAATCGAGAA 2853
Qy 217 AspValThrAspProArgValThrAspProProValAspAspGluArgLysAlaSer 236
Db 2854 -----ATCATGCTG--- 2862
Qy 237 GlyLysAspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCys 256
Db 2863 -----TCATCTCCGAAGATCCTACCTGTCTCAAGAC----- 2892
Qy 257 LysSerSerGlyThrAspGlyThrLeuAspGlyLysAsnAlaPheGlyLeuTyrSerMet 276
Db 2893 -----AATGATGCTGTG----- 2904
Qy 277 GlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAsp----- 292
Db 2905 -----CTGCTGAAAATAGTACAGCATCCAGTTCTCATGAAACATCA 2946
Qy 293 -----TrpGlyLysIleProAspGluSer-----GlnArgArgLeu 304
Db 2947 GTTGATGACACTTGGCAAGAGCTAGGGCGAGAAAGCCTTCAGGGGCACACAGAAAGGCTCT 3006
Qy 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
Db 305 HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
```

```

Db 3007 CACACCAACTCAAGCCAGAGCTGTCAAAGTGTATGAATTTGGAAATGCCGGAGACT 3066
QY 325 CyeThrSerMetValGluAsnPhleLeuGluTyrValileGluAsnHisArgIleTyrGlu 344
Db 3067 GTCAATGTTTATGTGAGGAATTAATGATTAATTTGAAGAAATCATCAA----- 3117
QY 345 GluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuLeu 364
Db 3118 -----TGTGGCGTGTCTCTCTC 3138
QY 365 VallysGlnIleLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPhe 384
Db 3139 TTAAGCATTAAGATCAATCAAGAGAGTCACTTGCCTTCTGTGAAACTTGAACATAT 3198
QY 385 LysCysSerArgValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeu 404
Db 3199 GAGTGGCATGAAGTGTCTGTAATATGTTATGAGAAGCTAAGGATCTCAAGAAAAAG 3258
QY 405 PheLeuGluHisThr-----GlnGlyLeuGlnPheAspCysPheGlyThr 419
Db 3259 TTTTCGGGTAGACCAAGTGAACAAGCAAGCAGACGACCAATCTACTCCAGTAATAACACA 3318
QY 420 AsnSerLysGlnSerValValSerThrLysLeuValAsnGluSer-----LeuSerGly 437
Db 3319 TCATCTACAAACAAGACACCTCAGCAAAATGAGAAGTGACCGGATCAATTTGTCATCAG 3378
QY 438 AlaThrValArgAspGluLysIleAsnThrLysSerMetArgAsnSerSerGluAsp--- 456
Db 3379 GTAACGACAATTCATGTTGATTTGGAAATGTTTCATGACGAGGCTCCACATGATATC 3438
QY 457 -----GluGluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAsp 474
Db 3439 TTGACCGGAGAGATGATATTAGAACAGAG----- 3468
QY 475 IleGluLysThrIleSerGlyIleLysLysLysLysLysGlnVal-----GlnLys 492
Db 3469 -----GAATGATATCTGTTCTAGAAATCTAGAGAGAAACATGTTTAAAGGATGAA 3522
QY 493 LeuValGlnGluHisGluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLys 512
Db 3523 CTTCTTGAAGAATCACAGAAAGAAAGATTAATTAATTAACATGTTTCTCTTAAGA 3582
QY 513 GlnLys-----LeuGluThrSerLysSer 520
Db 3583 GAAAGAATATTTCAGGATAAACAAGAAACGACACACACTGTTAGACATGCACAAGCAA 3642
QY 521 ValGluAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu 540
Db 3643 AAGGAAGTGCAGAAAGCTGCAGAAACATGCAAT----- 3675
QY 541 LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSerGluLysAsnGlu 560
Db 3676 TTGGTGTGGAACATCTTCGTAAAGTCACTTGTAT-----TCAGAGGACAGGGATGCC 3729
QY 561 CysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLysLeuAlaGluAspGlu 580
Db 3730 ACGGTGAAGCTAATAATTAATGATGTTTCTGCTGATGCAATTTTGAATCATCATG 3789
QY 581 AlaCysTrpIleAsnArgIleLys-----SerTrpAlaAlaLysLysLys 595
Db 3790 AGATGCAGCATAAACAAGTTGAAGTGCAGCAATCAACTTCATGGAACAAGGAGTTACAA 3849
QY 596 ValCys---ValProIleGlnSerGlyAsnAsnLysHisPheSerGlySerSerAsnIle 614
Db 3850 TTGAAGGAATTTCTCTCAGCAAGCAAAATCTGGCCACTTAGATCGTACGTTGATCAA 3909
QY 615 SerGlnAsnAlaProAsp 620
Db 3910 CAAATTCCTTTACCAGAT 3927

```

RESULT 4

US-10-425-115-159032

; Sequence 159032, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(5322)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 159032

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(1197)

; OTHER INFORMATION: unsure at all n locations

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_76611C.1

; US-10-425-115-159032

Alignment Scores:

Pred. No.:	7,94e-25	Length:	1197
Score:	336.50	Matches:	102
Percent Similarity:	48.63%	Conservative:	58
Best Local Similarity:	31.00%	Mismatches:	120
Query Match:	10.46%	Indels:	49
DB:	20	Gaps:	12

SEQ3_K705_COPY_650_1270 (1-621) x US-10-425-115-159032 (1-1197)

```

QY 3 AspAlaPheIleuPheGlySerLeuAsnProSerHisAspValLysHisValGlu 22
Db 142 GATGCCATCATATATTTGTAGTACTGGAATCCCATGAACTCCATGAGCCCTTCAG 201
QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db 202 AGGTCAGCATGAGTGCACAGTCTGAACTGTGCCTATTTTCGTTGTACTGCTTTT 261
QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
Db 262 ACAGTGCAGGAAAGCTCTTATTTTGCAAAAGCATGACCATATTTTGGATAGTAATAT 321
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 322 GTGAATATAACCCCTGAGCCATTTGCTGCTCAGTTGGGGCGCATCATTTCTCTTC 381
QY 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
Db 382 ATAGACTTGAGGAGTTGCAGCAACATAGCTAC-----TCAATGCTCTGCT 429
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
Db 430 GATGAGCTTTTCATGCAATAATGTAGATTTGGAGTTTTCACAAAATTTACTCAGCAAGGTT 489
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
Db 490 GAACTCAGAACTGCAAGCAATACAGCCATATCAAGCTTATCTGTGTGATCTTTC 549
QY 141 TyrSerSerAspSerThrLeuPheGlyGlu---AspHisIleLysLeuSerAspGluGlu 159
Db 550 TATTTCTAGGGCTATTGTTGTAGCGGTGAGAGGGAGGAATTCATCAGTTGATGTTGAC 609
QY 160 SerProAsnIle-----PheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLys 177
Db 610 CTGCCAAAGTCTCGGCCCTACTGGGTAAAGTTTACTTAATGGGAGGTCTCTCAGTGGCAG 669
QY 178 TyrProSerAspThrProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGlu 197
Db 670 TATATATCTGAGCCAGTACAAAGAACCGCTAGAAAGATA-----AATAACATGGAA 720

```


Query Match: 5.16% Indels: 283
DB: 10 Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x US-09-814-353-21548 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
DB 3342 GCGGACACAGCTGCTCATCTTTGACTCTGACTGAACCCCGAGATGACTTGCAGGACAA 3401

QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
DB 3402 GCCCGAGCGCATAGATTGCTCAGAGACGACGTAATAATTTACCGCTTAGTTACAAAG 3461

QY 42 CysThrValGluGluLysAla-----Leu 49
DB 3462 GGGACTGTGGAGGAGGAGATCATAGAACGGGCCAAAGAAAGATGGTATTAGATCATCTG 3521

QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
DB 3522 GTGATTCAGCGCATGACACCACTGCGCGGACGATCCTGGAAACAACTCAGGAAGTCC 3581

QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
DB 3582 AACTCAATCCTTTTATAAAGAGAGACTCAGAGCTATTGTGAATTTGGACGAGGAT 3641

QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
DB 3642 CTCTTCAAGAACTGGAA-----GGGGAGGAATCA 3671

QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
DB 3672 GAACCTCAGAAATGGAT-----ATAGATGAATTTTGGCGTTGGCTGAA 3716

QY 121 GlyGluGlnAsnGluValLysLeuCysLeuLeuLeuGluAlaLysHisAlaGlnGlyThr 140
DB 3717 ACGAGAGAGATGAAGTGTCAACAGTGCACAGATGAACCTCTA---TCACAGTTTAAG 3773

QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
DB 3774 GTTGCCAACTTTGCCAACTTGAAGATGAAGAGAGCTA-----GAAGAGCGT 3821

QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
DB 3822 CCTCAAGACTGGGATGAGATCATT-----3848

QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
DB 3849 -----CCAGAGAACAAAGAAAGTAAGAGAGAGAGCGGCGAGAGGCTAGAA 3902

QY 195 -----GlySerGluAlaSerProLysThrGly 203
DB 3903 GAAATTTATATGCTGCTCGAATTCGGAGTTCCACTTAAAGAGGCTCAGACAAATGACAGT 3962

QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspValThrAspProArgVal 223
DB 3963 GACTCTGACACTGAGTCTAAGAGCGGAGCCGACAGATCCTCTGCTTCTCAGAGTGAACG 4022

QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
DB 4023 GAAGAC-----TCTGATGATGACAAAG-----4046

QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
DB 4047 -----CCAAAGCGCAGAGGCGCTCGAGGAGTGTGCGGAAGGACCTCGTGGAGGGA 4097

QY 264 ThrLeuAspGlyAsn-----AspAlaPheCysLeuTyrSerMetGlySer 278
DB 4098 TTACTGATGCAGAGATCCGAGGTTTCATCAAGGGCTTATAAGAAAGTTT-----4145

QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
DB 4146 -----GGTCTCCCTTTTGAACGGCTGGAGTGCATGACGCTGAT---GCTGAGCTG 4193

QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
DB 4194 GTAGATAAGTCGGTGGCAGATCTG-----AAGCCCTGGGT 4229

QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
DB 4230 GAACTGATCCAC-----AACAGCTGTGTCTAGCAATCAGGAATACGAAGACGAGCTG 4283

QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
DB 4284 AAAGAAATGCCAGCGAGGAAAGGACCCAGGAAAGG---AGAGTCCCAACATCAAG 4340

QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
DB 4341 ATATCCGAGATT---CAGGTTAATGTGAATCCATTATCCAACTCAAGAGGAGTTT---4394

QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
DB 4395 -----GAGATGCTGCATTAATCTATCCCTGTGGACCTGAAGAAANAAAAA 4442

QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
DB 4443 TACTGCTTAACCTGTGCTCAAGCTGCACATTTTGTATGTAGAGTGGGGTGGAGAT 4502

QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
DB 4503 GATTCCTCGCTGTTGCTGGGATTTATGAACATGGCTATGGAACCTGGGAGTTAATTAA 4562

QY 422 sGlnSerValSerThrLysLeuValAsn-----432
DB 4563 ACAG-----ACCCAGAGCTTAATTAATTAACGTGACAAATTTCTGCGGTGGAGACA 4610

QY 433 ----GluSerLeuSerGlyValAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
DB 4611 GATAAAAGCCTCAGGGGAAGCAGCTACAGACCGGCGGATTACTTGTGTGAAGTGCCT 4670

QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
DB 4671 AGAAAGGGTCTCGAGAAAGAGGGGGCTGTGCACAGGTGGGGAAGAGGCCAAATTAAGAAG 4730

QY 464 Arg-----464
DB 4731 CGAAGCCTCGGTGAAGAAAGAAACAAAGTCCAGCTGAAAGAGGAGCATGGAAAT 4790

QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
DB 4791 GAGCTTTTCATCTCCTAGGCATTCAGATAATCCATCAGAGAGGAGAGTGAAGATGAT 4850

QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 4851 GGCCTTGGAAAAAAGTCCATGAAAAAAGAAACAGAAAG---AAGAAAGAGAAACAGGAGAAC 4907

QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----511
DB 4908 AAGGAGAAACAAATGAGTTCTAGGAAAGACAAAGAGGGGCAAGAGAAAGAAAGATCA 4967

QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----523
DB 4968 AAAGATAGAAGAGAGAGCGCTAAAGTGTGTGATGCCAATCTTCGAGTAATCAAGCGCA 5027

QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----539
DB 5028 TCTCAGGGTCTGCTCCATATTACAGCAGGAAGTGAACCTGTCCCATTCGAGGAGGATGAG 5087

QY 539 -----539
DB 5088 GATGATGATCTCGACACGAGACATTCACATATGTAAGAGAGAGATGAGGCCGTGAAA 5147

QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
DB 5148 AAGGCACCTGAAACAGCTGCACAAACCTGCAAGGGGCTCAACGTGCAAGAACAGCTGGAA 5207

QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer-----564


```
QY 395 TyrSerIleLeuSerCysMetIysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
DB 4443 TACTGCTTAACCTGCTGCTCAAGCTGCACATTTTGTAGTAGTGGGGTGGGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
DB 4503 GATCTCGGCTGTGCTGGGATTTATGAACATGGCTATGGAACTGGGAGTTAATAAA 4562
QY 422 sGlnSerValSerThrLysLeuValAsn-----
DB 4563 ACAG-----ACCAGAGCTTAATTAACCTGACAAAATTCGCGGTGGAGACA 4610
QY 433 ----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn--ThrLysSerMet 450
DB 4611 GATAAAAGCCTCAGGGGAAGCAGCTACAGCCGAGCGATTTACTTGTGAAGCTGCTC 4670
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
DB 4671 AGAAGGGTCTGGAGAAGAGGGGGCTGTGACAGTGGGGAGAGGCCAAATTTAAGAAG 4730
QY 464 Arg-----
DB 4731 CGGAAGCCTCGGTAAAGAAAGAAACAAGTCCCGAGCTGAAGAGGAGCATGAATT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
DB 4791 GAGCTTTTCATCTCCTAGGCATTCAGATAATCCATCAGAAGAGGGAGAAAGTGAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 4851 GCCTTGGAAAGTCCAAATGAAAGAAACACAGAG--AAGAAAGAGAAACAAGGAGAAC 4907
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----
DB 4908 AAGGAGAAACAATGATGTTCTAGGAAGACAAGAGAGGGGACAGGAAGAAGAGTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----
DB 4968 AAGAGATAAGAAAGAGAGAGCTAAAGTGGTGTGATGCCAATCTTCAGATAATCAAAGCGA 5027
QY 524 ----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----
DB 5028 TCTCAGGGTCTGCTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGGAGATGAG 5087
QY 539 -----
DB 5088 GATGATGATCTGGACCGAGACATTCACCATATGTAGGAGAGAGATGAGGCCGTGAAA 5147
QY 540 ----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
DB 5148 AAGGCACTGAACAGCTGCACAAACCTGACAAGGGGCTCAACGTGCAAGAACAGCTGGAA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer-----
DB 5208 CACACCCGGAAGTGGCTGCTGAAATTCGGAGACCGGATAGCCGAGTGGCTTAAGCCTAC 5267
QY 564 -----
DB 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGAACCTATGATTTTGTTCAGATT 5327
QY 565 ----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
DB 5328 ACAGAAATTTGATGCTCGAAAACCTGCATAAGTTATACAGATGGCTCATAGAAGAAAGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
DB 5388 CAAGAAGAGAG-----GAGCAAAAGAAAGAAAGACCGAC 5420
QY 598 ValProIleGlnSerClyAsnAsnLysHisPhe-----SerGlySerSerAsn 613
DB 5421 GTG-----ACTGGGGGTAAAGAAACCATTTTCGTCCAGAGGCCTCAGGCCTCCAGCCGG 5471
```

```
QY 614 -----IleSerGlnAsn 617
DB 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 8
US-10-651-237-27
; Sequence 27, Application US/10651237
; Publication NO. US20050048494A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: ADS-5003 US NP
; CURRENT APPLICATION NUMBER: US/10/651,237
; CURRENT FILING DATE: 2003-08-27
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
US-10-651-237-27
Alignment Scores:
Pred. No.: 1,098-05 Length: 7764
Score: 166.00 Matches: 162
Percent Similarity: 33.99% Conservative: 113
Best Local Similarity: 20.02% Mismatches: 251
Query Match: 5.16% Indels: 283
DB: 21 Gaps: 37
SEQ3_K705_COPY_650_1270 (1-621) x US-10-651-237-27 (1-7764)
QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisVal 21
DB 3342 GGGGACACAGCTGCTATCTTCTGACTCTGACTGACACCCCGACAGATGCTTGACGACCAA 3401
QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
DB 3402 GCCCGAGCGCATAGAAATTTGTCAGAAAGACGAGGTAAATATTTACCGCTTAGTTACAAAG 3461
QY 42 CysThrValGluGluLysAla-----Leu 49
DB 3462 GGGAGCTGGGAGGAGGAGATCATAGACGGCGGCAAAAGAGATGGTATTAGATCATCTG 3521
QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
DB 3522 GTGATTCAGCGCATGACACCACTGGCGGAGCATCTGGAAAAACAACACTCAGGAAGGTCC 3581
QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
DB 3582 AACTCAAATCTCTTTTAAATAAGACAGAGCTGACAGCTATTTTGAATTTTGACGACAGGAT 3641
QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
DB 3642 CTCTTCAAAGAACTCGAA-----GGGAGGGAATCA 3671
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGly 120
DB 3672 GAACCTCAGGAATGGAT-----ATAGATGAATTTTGGGTTGGCTGAA 3716
QY 121 GlyGluGluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
DB 3717 ACGAGAGAGATGAAGTGTCAACAAGTGCACAGATGAATCTCTA--TCACAGTTTAAG 3773
QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluGluSer 160
DB 3774 GTTCCCAACTTTGCAACAATGCAAGATGAAGAAGAGCTA-----GAAGAGCGT 3821
QY 161 ProAsnIlePheTrpSerLysLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
DB 3822 CCTCAAGGACTGGGATGAGATCAIT-----3848
```

QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu----- 194
DB 3849 -----CCAGAGGAACAAAGGAAAAAGTAGAGGAGGAAGCGCGCAGAGGAGCTAGAA 3902
QY 195 -----GlySerGluAlaSerProLysThrGly 203
DB 3903 GAAATTTATATGCTGCTCGAATTCCGAGTTCCACTAAAGGCTCAGACAAATCAGAGT 3962
QY 204 AspGlyGlyAsnAlaLysArgLysLysLysAlaSerAspAspValThrAspProArgVal 223
DB 3963 GACTCTGACACTAGTCTTAAGAGGCGAGCCAGAGATCCTCTGCTTCTGAGAGTGAACG 4022
QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
DB 4023 GAAGAC-----TCTGATGATGACAGAAG----- 4046
QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
DB 4047 -----CCAAAGCGAGGGGCTCGAGAGTGTGCGGAAGACCTCTGTGGAGGA 4097
QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
DB 4098 TTCTACTGATCGAGATCCGAAGTTTCATCAAGGCTTATAAGAAGTTT----- 4145
QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
DB 4146 -----GGTCTCCCTCTTGAACGGCTGGAGTGTCTTAGCACGCTGAT---GCTGAGCTG 4193
QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCys 316
DB 4194 GTAGATATAGTCGGTGGCAGATCTG-----AAGCGCTGGGT 4229
QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
DB 4230 GAACATGATCCAC-----AACAGTGTGTGTCAGCAATGCAGGAATATGAAGAGCAGCTG 4283
QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAla 356
DB 4284 AAGAAAAATGCCAGCGAGGGAAGAACCCAGGGAAGAGG---AGAGGTCCACAAATCAAG 4340
QY 357 LeuSerTrpIleAlaAlaLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
DB 4341 ATATCCGAGTT---CAGTTTAATGTGAATTCATTCATTCACATGAAGAGAGTTT--- 4394
QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
DB 4395 -----GAGATGCTGATAAATCTATCCCTGTGGACCTGGAAGAAAAA 4442
QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
DB 4443 TACTGCTTAACTGCTGCTGTCGAAGCTGCACATTTTGTATGATGAGTGGGGGTGGAAGAT 4502
QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
DB 4503 GATTCCTCGCTTCTGCTGGGATTTATGAACTGCTATGGAATCTGGAGTTAATTA 4562
QY 422 sGlnSerValValSerThrLysLeuValAsn----- 432
DB 4563 ACAG-----ACCCAGAGCTTAAATTAATTAAGTAACTGCAAAATTTCTGCGGTGGAGACA 4610
QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
DB 4611 GATAAAAGCCCTCAGGGGAAGCAGCTACAGACCCGAGCGGATTAATTTGTTGAAGCTGCTC 4670
QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
DB 4671 AGAAGGGTCTCGAAGAGAGGGGCTGTGACAGGTGGGAGGAGGCCAAATTAAGAAG 4730
QY 464 Arg----- 464
DB 4731 CGGAAGCCTCGGGTAAAGAGGAAACAAAGTGCCAGGCTGAAAGAGGAGCATGGAATT 4790
QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480

DB 4791 GAGCTTTCATCTCTAGGCATTTCCAGATAATCCATCAGAAGAGGGGAGAGTGAAGATGAT 4850
QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGluHisGluGluLys 500
DB 4851 GGCITGGAAAAAGTCCAATGNAAAAAAACAAGAG---AAGAAAGAGACAAGGAGAAC 4907
QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys----- 511
DB 4908 AAGGAGAAACAATGAGTTCTAGGAAAGACAAAGAGGGGACAAGGAAAGAAAGAGTCA 4967
QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla----- 523
DB 4968 AAGATAAGAAAGAGAACCTAAAGTGTGATGTCCTCAAAATCTTCAGGTAAATCAAAGCGA 5027
QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp--- 539
DB 5028 TCTCAGGGTCTGTCCATATTACAGCAGGAAGTGAACCTGTCCCATTTGGAGAGGATGAG 5087
QY 539 ----- 539
DB 5088 GATGATGATCTGGACCGAGACATTCAGCATATGTAAAGGAGAGGATGAGGCCCGTGAAA 5147
QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552
DB 5148 AAGCAGCTGAAACAGCTCGACAAACCTGACAAGGGGCTCAACGTCGAACAGAGCTGGAA 5207
QY 553 -----GluIleLysSerGluLysAsnGluCysLeuLysSer--- 564
DB 5208 CACACCCGNACTGCCTGCTGAAATCGAGACCGGATAGCCGAGTGCCTTTAAAGCCTAC 5267
QY 564 ----- 564
DB 5268 TCAGATCAGGAGCACATCAAACTCTGGAGGAGGAACCTATGGATTTTGTTCCTCAAGTTT 5327
QY 565 -----LeuGluGlnMetHisGluValAlaLysLysLysLeuAla 577
DB 5328 ACAGAAATTTGATGCTCGAAAACTGCATAAGTTATACAAGATGGCTCATAGAAGAGGTCT 5387
QY 578 GluAspGluAlaCysTrpIleAsnArgIleLysSerTrpAlaAlaLysLeuLysValCys 597
DB 5388 CAAGAAGAAGAG-----GAGCAAAAGAGAAAGACGAC 5420
QY 598 ValProIleGlnSerGlyAsnAsnLysHisPhe-----SerGlySerSerSerAsn 613
DB 5421 GTG-----ACTGGGGGTAAAGAAACCATTTTCCTCCAGAGGCGCTCAGGCTCCAGCCGG 5471
QY 614 -----IleSerGlnAsn 617
DB 5472 GACTCTCTGATATCTCAGTCC 5492
RESULT 9
US-10-782-413-27
; Sequence 27, Application US/10782413
; Publication No. US20050048526A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: VDX-5002 CIP
; CURRENT APPLICATION NUMBER: US/10/782,413
; PRIOR FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: 10/651,237
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
US-10-782-413-27
Alignment Scores:

Pred. No.: 1.09e-05 Length: 7764
 Score: 166.00 Matches: 162
 Percent Similarity: 33.99% Conservative: 113
 Best Local Similarity: 20.02% Mismatches: 251
 Query Match: 5.16% Indels: 283
 DB: 21 Gaps: 37

SEQ3_K705_COPY_650_1270 (1-621) x US-10-782-413-27 (1-7764)

QY 2 AlaAspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValHisVal 21
 DB 3342 GCGGACACAGCTGCTCATCTTTGACTCTGACTGGAAACCCAGCAATGACTTGCAGGCACAA 3401
 QY 22 GluLysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerVal 41
 DB 3402 GCCCGAGCGCATAGAAATTTGTCAGAAAGACGAGTAATATTACCGCTTAGTACAAAG 3461
 QY 42 CysThrValGluGluLysAla-----Leu 49
 DB 3462 GGGACTGTGGAGGAGGATCATAGAACGGGCCAAAGAAAGATGGTATTAGATCATCTG 3521
 QY 50 IleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsn-----64
 DB 3522 GTGATTGAGCGCATGCACACCACTGCGCGGACGATCTCTGGAACAACAACACTCAGGAAGTCC 3581
 QY 65 -----LeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyr 80
 DB 3582 AACTCAAAATCCTTTTAATAAAGAGAGAGCTGCACAGCTATTATTGAATTTTGAGCAGAGGAT 3641
 QY 81 LeuPheAspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPhe 100
 DB 3642 CTCTTCAAGAACTGGAA-----GGGAGGAGATCA 3671
 QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerLysGly 120
 DB 3672 GAACCTCAGAAATGGAT-----ATAGATGAATTTTTCGGTTGGCTGAA 3716
 QY 121 GlyGluGluAsnGluValLysLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThr 140
 DB 3717 ACAGAGAGAATGAAGTGTCAACAGTGCACAGATGAACCTCTA---TCACAGTTTAAG 3773
 QY 141 TyrSerSerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGluSer 160
 DB 3774 GTTGCCCAACTTTCACAACTGGAAAGATGAAGAAGCTA-----GAAGAGCGT 3821
 QY 161 ProAsnIlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSer 180
 DB 3822 CCTCACAAGGACTGGGTGATGATCAT-----3848
 QY 181 AspThrProGlnArgAsnArgLysArgValGlnTyrPheGlu-----194
 DB 3849 -----CCAGAGAACAAAGAAAGTAAGTAGAGGAGGAGCGGCGAGAGGAGCTAGAA 3902
 QY 195 -----GlySerGluAlaSerProLysThrGly 203
 DB 3903 GAAATTTATATGCTCGCTCGAATTCGAGGTTCACATAAAGAGCTCAGACAAATGACAGT 3962
 QY 204 AspGlyGlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgVal 223
 DB 3963 GACTCTGACACTAGTCTAAGAGCGAGGCCAGACAGATCCTCTCTCTGAGAGTGAAACG 4022
 QY 224 ThrAspProValAspAspGluArgLysAlaSerGlyLysAspHisMetGlyAla 243
 DB 4023 GAAGAC-----TCTGATGATGACAAGAA-----4046
 QY 244 LeuGluSerProLysValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGly 263
 DB 4047 -----CCAAAGCGCAGAGGCGCTCCGAGAGGTGTCGCGAAGGACCTCTCGGAGGGA 4097
 QY 264 ThrLeuAspGlyAsn-----AspAlaPheGlyLeuTyrSerMetGlySer 278
 DB 4098 TTTACTGATGCAGATCCGAAGTTTCATCAAGGCTTATAAGAGTTT-----4145

QY 279 HisIleSerGlyIleProGluAspMetLeu-----AlaSerGlnAspTrpGlyLysIle 296
 DB 4146 -----GGTCTCCCTCTTGAACGCTGGAGTCTGTAGCAGCTGAT---GCTGAGCTG 4193
 QY 297 ProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaValLeuCys 316
 DB 4194 GTAGATAGTCTGGTGCAGATCTG-----AAGCCCTGGGT 4229
 QY 317 GlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyrVal 336
 DB 4230 GAACATGATCCAC-----AACAGCTGTGTGTGAGCAATGTCAGCAATATGAAGCAGCTG 4283
 QY 337 IleGluAsnHisArgIleTyrGluGluProAlaThrPheGlnAlaPheGlnIleAla 356
 DB 4284 AAGAAATATCCACGAGGAGGAAAGGAGCCAGGGAAGG---AGAGTCCCAACAATCAAG 4340
 QY 357 LeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuVal 376
 DB 4341 ATATCCGGAGTT---CAGGTTAATGTGAATTCATTTATCCACATGAGAGGAGTT---4394
 QY 377 ArgAlaAsnSerGluLeuAlaPheLysCysSerArgVal-----GluValAspTyrIle 394
 DB 4395 -----GAGATGCTGCATAAATCTATCCCTGTGGACCTGAAGAAAAA 4442
 QY 395 TyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPhe 414
 DB 4443 TACTGTCTAACCTGTCTGTCAAAGCTGCACATTTTGTAGTGTAGAGTGGGGTGGAGAT 4502
 QY 415 AspCys-----PheGly-ThrAsnSer-----Lys 422
 DB 4503 GATCTCGCTGTGTCTGGGATTTATGAACATGGCTGTGAAACTGGGAGTTAATTA 4562
 QY 422 sGlnSerValSerThrLysLeuValAsn-----432
 DB 4563 ACAG-----ACCAGAGCTTAAATTAACATGACAAATTTCTCGCGTGGAGACA 4610
 QY 433 -----GluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn---ThrLysSerMet 450
 DB 4611 GATAAAAGCCTCAGGGGAAGCAGCTACAGACCGAGCGGATTAATTTGTTGAAGCTGCTC 4670
 QY 451 ArgAsnSerSerGlu-----AspGluGluCysMetThrGluLys 463
 DB 4671 AGAAGGGTCTGGAGAAGAGGGGGCTGTGACAGTGGGAGGAGGCCAATTAAGAAG 4730
 QY 464 Arg-----464
 DB 4731 CGGAAGCCTCGGTAAAGAGGAACAAAGTCCCGAGCTGAAGAGGAGCATGGAAT 4790
 QY 465 -----CysSerHisTyrSerThrAlaThrArgAspIleGluLysThrIleSer 480
 DB 4791 GAGCTTTTCATCTCTAGGCAATTCAGATAATCCATCAGAGAGGGAGAGTGAAGATGAT 4850
 QY 481 GlyIleLysLysLysTyrLysLysGlnValGlnLysLeuValGlnGlnHisGluGluLys 500
 DB 4851 GGGTTGGAAAAAGTCCCATGTAATAAAAAACAGAG---AAGAAAGAGAACAGGAGAAC 4907
 QY 501 LysMetGluLeuLeuAsnMetTyrAlaAspLys-----511
 DB 4908 AAGGAGAAACAAATGAGTTCTAGGAAGACAAAGAGGGGAGCAAGGAAGAAAGAGTCA 4967
 QY 512 LysGlnLysLeuGluThrSerLysSerValGluAla-----523
 DB 4968 AAGATAAAGAAAGAGAGGAGCTTAAAGTGGTGTGCCAAATCTTCGAGTAAATCAAGCGA 5027
 QY 524 -----AlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAsp-----539
 DB 5028 TCTCAGGGTCTGTCTCATATTACAGCGAAGTGAACCTGTCCCTTCCAGAGGAGTGA 5087
 QY 539 -----539
 DB 5088 GATGATGATCTGGACCGAGACATTCAGCATATGTAGGAGAGGATGAGCCCGTGA 5147
 QY 540 -----LeuLysLeuLeuAspHis-----AsnTyrGluArgLysPheAsp 552

Db	5148	AAGGC	ACTGAAACAGCTCGACAAACCTGCAAGGGGCTCAACGTGCAAGAACACAGCTGGAA	5207
Qy	553	-----	-----GluIleLysSerCluLysAenGluCysLeuLysSer---	564
		-----	-----	
		-----	-----	
		-----	-----	
Db	5208	CACACCGGAACTGCCTGCTGGAATACTGGAGACCGGATAGCCGAGTGCCTTAAGCCTAC	5267	
		-----	-----	
		-----	-----	
Qy	564	-----	-----	564
Db	5268	TCAGATCAGGAGCACATCAAACCTCTGGAGGAGGAACCTATGGATTTTGTTCCAAGTTT	5327	
		-----	-----	
		-----	-----	
Qy	565	-----	-----LeuGluGlnMetHisGluValalalysLysLysLeuAla	577
		-----	-----	
		-----	-----	
		-----	-----	
Db	5328	ACAGAATTTGATGCTCGAAAACCTGCATAAGTTATACAAGATGGCTCNTAGAAAAGGTCT	5387	
		-----	-----	
		-----	-----	
		-----	-----	
Qy	578	GluaspluAlaCysTrpIleAsnArgIleLysSerTrpAlaLalalysLeuLysValCys	597	
		-----	-----	
		-----	-----	
Db	5388	CAAGAGAAGAG-----	-GAGCAAAAGAGAAGAACGAC	5420
		-----	-----	
		-----	-----	
Qy	598	ValProIleGlnSerGlyAsnAenLysHiePhe-----	-SerGlySerSerAsn	613
		-----	-----	
		-----	-----	
		-----	-----	
Db	5421	GTG-----	-ACTGGGGGTAGAAAACCATTTGTCCTCAGAGGCTTCAGGCTCCAGGCGG	5471
		-----	-----	
		-----	-----	
Qy	614	-----	-IleSerGlnAsn	617
		-----	-----	
		-----	-----	
Db	5472	GACTCTCTGATATCTCAGTCC	5492	

RESULT 10
 US-10-369-493-26897
 ; Sequence 26897, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 26897
 ; LENGTH: 3045
 ; TYPE: DNA
 ; ORGANISM: Neurospora crassa
 ; US-10-369-493-26897

Alignment Scores:	
Pred. No.:	4.25e-06
Score:	164.00
Matches:	131
Conservative:	112
Best Local Similarity:	37.16%
Best Local Similarity:	20.03%
Query Match:	5.10%
DB:	17
Gaps:	29
Indels:	176
Mismatches:	236
Length:	3045

SEQ3 K705 COPY 650 1270 (1-621) x US-10-369-493-26897 (1-3045)

Qy	17	AspValLysHisValGluLysIleLysIleGluSerCysSerGluArgThrLysIlePhe	36
Db	673	GACATCGAGCATTGGCGCAAGTGCACACCGATCTCAAAAGAAGAGGAGATTCTCTATCT	732
Qy	37	ArgLeuTyrSerCysThrValGluLysAlaLeuIleuAlaArgGlnAsnLys	56
Db	733	CGCTTG-----CAGGAAGAAGCT-----ACTCGCAACACCGAG	765
Qy	57	ArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrp	76
Db	766	CAGGCCACCCAGATGTCGGATTCCTTTAAACCGAAGCTGAAC-----	807

Qy	77	GlyAlaSerTyrLeuPheAspIysLeuAspHisPheHisSerSerGluThrProAspSer	96
Db	808	-----GATATCATGGTATCCGCGCAGCGTTCGCGTTCATCAAA	849
Qy	97	GlyValSerPheGluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeu	116
Db	850	GGCGGCGCCGACCTCATGCATAGCGGCACCAATCCTTGAAGTCTTCGATAACATATTG	909
Qy	117	SerSerLysGlyGlyGluGluAenGluValLysLeuCysLeuLeuLeuGluAlaLys---	135
Db	910	AACCTGTAC-----GAAGACACCGAGCAACGACTTGTTCAAACCCGAAGAAGCAAGGAC	963
Qy	136	HisAlaGlnClyThrTyrSerSerAspSerThrLeuPheGlyGluAspHisIle---	154
Db	964	AAGGCACAGCAGCGCCGAAGGAATCCGCCACACAGATCCCGCAGTGGAAACAATCCAGAA	1023
Qy	155	LeuSerAspGluGluSer-----	162
Db	1024	CTCACGTCTGAAAACAACAACGCTTAAAGAACAGAGAGTCAAGCAAAAAGAAAGCCGAC	1083
Qy	163	IlePheTrpSerLysLeuLeuGlyGlyLysAsnProMetTrpLysTyrProSerAspThr	182
Db	1084	GCCTTTTACAGAAACTC-----AGAAAAAGGATGGCTTGAGCCGACCGAGGAT	1134
Qy	183	Pro-----GlnArgAsnArgLysArgValGlnTyrPheGluGly	195
Db	1135	CCATCGCGCTCACTGCCCTCGACAGAGTCTGTGAAGACCAACAAGATCGCTGGGAG	1194
Qy	196	SerGluAlaSerProLysThrGlyAspGlyGlyAsnAlaLysLysArg---	214
Db	1195	TCTGAAAGAAGATGCCCTCGAACCAAGAAATCACTAATCTCAACAATTCGATCGAAGAGCC	1254
Qy	215	SerAspAspValThrAspProArgVal-----ThrAspProProValAspAsp	230
Db	1255	AATGTGGCGCTCAAGGACATGAAGCTTATCCAGCTGGAGAGGATCTCTGGTTCGATCTA	1314
Qy	231	AspGluArgLysAlaSerGlyLysAspHisMetGlyAlaLeuGluSer-----	248
Db	1315	CAACAACGCTCAGGAA-----GGGGATATCAAAAGCTTGAACCAACACAG	1356
Qy	249	ValIleThrLeuGlnSerSerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsn	268
Db	1357	GTTCGTGACCTCAAGCAAAAATTTGGCGAGCAAGCCGACGCCGCCGCCGA-----	1409
Qy	269	AspAlaPheGlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeu	288
Db	1410	GAAGGATATTCAGTCTTTGGATTGGAGACAAGTCTTTTGGAGACCAAGCAACGCTGT	1469
Qy	289	AlaSerGlnAspTrp-----	295
Db	1470	GCAAGCCGAAAGAGGAGTCTTGCAGCAGCAGCTTTCTCAGACGAAGCAAGCTTGGAAAA	1529
Qy	296	IleProAspGluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeu	315
Db	1530	GGTGGAGACACCTCAAGAA-----CACGGCGTCGCAAAAGATGGATT-----	1574
Qy	316	CysGlnValLeuHisLeuSerAspAlaCysThrSerMetValGlyAsnPheLeuGluTyr	335
Db	1575	-----CGAGACTCAAAAGAAATGA	1592
Qy	336	ValIleGluAsn-HisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnI	355
Db	1593	GTGTCTCAAGGCAAGAAAGTCTCTGGAGGAAGAG-----	1626
Qy	355	eAlaLeuSerTrpIleAlaAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLe	375
Db	1627	-----ATCAGCAGGCTCAAGAGGAGGTTTGACTTGCCCAAGGAAGATTCCTCCG	1673
Qy	375	u-----ValArgAlaAsnSerGluLeuAlaPheLysCysSerArgVa	389
Db	1674	TGGCCCAACAANTCGTGGAGATCAACCGGAACCCCAACATTTGCCGAAGGATCACTCGCT	1733
Qy	389	IleuVal-----AspTyrIleTyr	395

```

Db      1734  CGAACAACTCATGACGAGCAAAACCTCACTACCGCTACCGCAGCGACCTTCGTCGA 1793
QY      :||||:
395  rSerIleLeuSerCysMetLysSerLeu-----PheLeuGluHisThrGlnG1 411
Db      :||||:
1794  AGGACTGGTTGCCAAGTTGAAGAGCGCGCAGGACACACAAGGACTTGGATCCAGTGAGGA 1853
QY      :||||:
411  yLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerThrLysLeuVa 431
Db      :||||:
1854  TGTCAATTCGTAATGCAATCGGC-----CTTCT 1880
QY      :||||:
431  lAsnGluSerLeuSerGlyAlaThrValArgAspGluLysIleAsn----- 446
Db      :||||:
1881  CAATCGTCTTGAAG---GAGCTGCTCTTGAAGAGACGACAGAGATGCTTCAGAC 1937
QY      :||||:
447  -----ThrLysSerMetArgAsnSerSerGluAspGluGlu-----Cy 459
Db      :||||:
1938  TTTGCAAAATGAGTGAATGAGCTCTGAGGAGGCGCAAGGAGGACCGCAACAGCGC 1997
QY      :||||:
459  sMetThrGluLysArgCysSerHisThrSerThrAlaThrArgAspIleGluLysThrI1 479
Db      :||||:
1998  GCAGTTGTCAGAGCATCATTCACAAATACGAAGGCGCAAGGAAAGGCTTGAGAAGGAGGT 2057
QY      :||||:
479  eSerGlyIleLysLysLysLysLysGlnValGlnLysLeuValGlnGluHisGluG1 499
Db      :||||:
2058  C-----AAGTTCTAAAGACGAGATCAAGAAGTTGAAGCAGGAGATCACTGA 2105
QY      :||||:
499  uLysLysMetGluLeuLeuAsn-----MetTyrAl 509
Db      :||||:
2106  GAAGAAACCTCACTGCGCAACAGCAGCAGAGAGAGATATGCTCAAGGAAAGCTACGA 2165
QY      :||||:
509  aAspLysLysGlnLysLeuGluThrSerLysSerValGluAlaValIleArgIleth 529
Db      :||||:
2166  AGAGCAAAATCAAAAATCTCAATCGGACCAACAAAGGCTGCTCGGAGCTGAAAGTCAA 2225
QY      :||||:
529  rCysSerArgThrSerThrGlnVal-----GlyAspLeuLysLeuLeuAspHisAs 546
Db      :||||:
2226  ACACCAAAACGAGCTGACCGAGCTCAGAAAGGACGCGGACTTGAAA-----GAAACCAA 2279
QY      :||||:
546  nTyrGluArgLysPheAspGluLysSerGluLysAsnGluCysLeuLysSerLeuG1 566
Db      :||||:
2280  CCTCTGCAAGATTAGACACTCAGGAGCAGATGAAGTGAGGAGGACAGGCTTCA 2339
QY      :||||:
566  uGlnMetHisGluValAlaLysLysLysLysLeuAlaGluAsp 579
Db      :||||:
2340  AGCCGATTACCGCTGAGAGGCGGAAGCTCAGGAAGGAT 2379

```

RESULT 11

```

US-10-425-115-145108
; Sequence 145108, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 145108
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1
US-10-425-115-145108

```

```

Alignment Scores:
Pred. No.:      8,61e-07      Length:      661
Score:          161.00      Matches:      50

```

```

Percent Similarity: 44.51%      Conservative: 23
Best Local Similarity: 30.49%      Mismatches: 75
Query Match: 5.01%      Indels: 16
DB: 20      Gaps: 5
SEQ3_K705_COPY_650_1270 (1-621) x US-10-425-115-145108 (1-661)

QY      239  AspHisMetGlyAlaLeuGluSerProLysValIleThrLeuGlnSerSerCysLysSer 258
Db      :||||:
6  GATGAAGTTGCTAAATCCCAAGCAATCAACTGTCTCCACAAAAACAAAAAAGTGTAAAGGA 65
QY      259  SerGly-----ThrAspGlyThrLeuAspGlyAsnAspAlaPhe 271
Db      :||||:
66  ATTCAATTCAGAAACTTCTGAGATTGCCAGATGGGTGTCATCTCGATTTTGATAATGAT 125
QY      272  GlyLeuTyrSerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGln 291
Db      :||||:
126  CATATGTGTCTGTAATAATCAT---GACCTGCTCTCTGAAATGATGATCTCCCAAGTGCT 182
QY      292  AspTyrGlyLysIleProAsp-----GluSerGlnArg-----ArgLeu 304
Db      :||||:
183  GACCTTCATCTACTCTCTGAGACGAGGTGGAACACATGAGGACACCTAAAGTAGTTTT 242
QY      305  HisThrValLeuLysProLysMetAlaLysLeuCysGlnValLeuHisLeuSerAspAla 324
Db      :||||:
243  CAGCTGAACTTAAGCGTGAGCTGTCAAGCTAAATACCGTGTAAAGCTTACCAGATAAT 302
QY      325  CysThrSerMetValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGlu 344
Db      :||||:
303  GTACTTTTCTGCCCAACAGCTCCCTCGAATATCTTTTAATATCATTTAGTTGCGCG 362
QY      345  GluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuLeu 364
Db      :||||:
363  GAGCACGCGGACATCTGCATGCGTTCAACATAGCCTTGTGTGGCGTGCTGCTCTTTT 422
QY      365  ValLysGlnIle---LeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAla 383
Db      :||||:
423  CTGAATACACTGAGTGGATCATCGAGAATCACTTGCCTTGCCTTCAGATGGAAT 482
QY      384  PheLysCysSer 387
Db      :||||:
483  TATGAATGCAAT 494

```

RESULT 12

```

US-10-956-157-4967
; Sequence 4967, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4967
; LENGTH: 5119
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4967

```

```

Alignment Scores:
Pred. No.:      3,14e-05      Length:      5119
Score:          159.00      Matches:      136
Percent Similarity: 39.97%      Conservative: 107
Best Local Similarity: 22.37%      Mismatches: 211
Query Match: 4.94%      Indels: 155
DB: 21      Gaps: 28
SEQ3_K705_COPY_650_1270 (1-621) x US-10-956-157-4967 (1-5119)

```


; Sequence 305, Application US/10843641A
; Publication No. US20050064454A1
; GENERAL INFORMATION:
; APPLICANT: Avalon Pharmaceuticals, Inc.
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets
; FILE REFERENCE: 689290-189
; CURRENT APPLICATION NUMBER: US/10/843,641A
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: US/09/873,367
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,832
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/964,824
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/967,768
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/09/968,007
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 847
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 305
; LENGTH: 5857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-843-641A-305

Alignment Scores:

Pred. No.:	3,85e-05	Length:	5857
Score:	159.00	Matches:	136
Percent Similarity:	39.97%	Conservatives:	107
Best Local Similarity:	22.37%	Mismatches:	211
Query Match:	4.94%	Indels:	155
DB:	21	Gaps:	28

SEQ3_K705_COPY_650_1270 (1-621) x US-10-843-641A-305 (1-5857)

QY	31	GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle	50
DB	1840	GAAGAACTCATCAGAAGGAGATAAGGCTCTGTATACCGCCACGGAAG-----	1890
QY	51	LeuAlaArgGluAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr	70
DB	1891	CTTTCCAAAGAGAACGATTCATTGAAAGCAAGCATGGACATGCCCAACAAAGAGAACTCA	1950
QY	71	HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisPheHisSer	90
DB	1951	GATGTGATAGCTTATGGAGTCC-----AAACTG-----	1980
QY	91	SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis	110
DB	1981	-----GAGACTGCCATCGCATCCACAGCGCGATGGAAGAACTGAAGGTA	2028
QY	111	GluPheSerSerIleLeuSerSerLysGlyGluGluAsnGluValLysLeu-----	128
DB	2029	TCTTTCCAGCAAGGGCTTGGACAGAGACGGCAGAAATTTGCTGAATCAAAACACAAATA	2088
QY	129	---CysLeuLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSerSer-AspSerThrLe	147
DB	2089	GAGAAATGAGACTAGATTACCAACGAAATAGAAAATTTGCAGAAATCAACAGACTCT	2148
QY	147	uPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSer---	166

Db	2149	GAACGGGCTGCCCATGCTAAAGAGA-----	2181
QY	167	-----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrProS	180
Db	2182	TTGAGGGCTAAACTGATGAAGTTATTAAAGAAAGGAAACACAGTCTGGAACCATCAGG	2241
QY	180	erAspThrProGln-----ArgAsnArgLysArgValGlnTyrP	193
Db	2242	TCGAAACTGGACAAAGCAGACAGCAGCATCTCGTAGAAATCGAAGACACACGTATAACAA	2301
QY	193	heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys	210
Db	2302	TTACAGGAAGCTGAAATAAAGGTAAGGAGCTAGAGGTACTCCAGCCCAATGCAATGAA	2361
QY	211	ArgLysLysAlaSerAspAspValThrAspPro---ArgValThrAspProProValAsp	229
Db	2362	CAAACCAAGTTATTGATAATTTTACATCAGCTCAAGGCTACTGAGAGAAAAGCTCTTG	2421
QY	230	AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu	245
Db	2422	GATCTTGATGCATCTCGGAAGCCAGTTCGGAAGGTAAATCGAATCGAAGAACTTAGA	2481
QY	246	SerPro-----LysValIleThrLeuGln-----Ser	254
Db	2482	CAGCAGCTTGAGCGAGCTGAGAAACAGATTAAACATTTAGAGATTGAAAGAAATGCTGAA	2541
QY	255	SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr	274
Db	2542	AGTACCAAGCTGTAGCATTTACAGAGAGCTCCAGGGGAGAGAG---CTAAAGCTTACT	2598
QY	275	SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly	294
Db	2599	AACCTTCAGAAAATTTGAGT-----GAAGT-CAGTCAAGTGAAGAGACCTTGA	2648
QY	295	LysIleProAspGluSerGlnArgArgLeu-----HisThrValLeu	308
Db	2649	AAAAGAACTTCAGATTTTGAAGAAAAGTTTGCTGAAGCTTCAGAGGAGCGACTCTGT	2708
QY	309	LysProLysMetAlaLysLeuCysGln-ValLeuHisLeuSerAspAlaCysThrSerMe	328
Db	2709	TCAGAGAGATGTCAGAAACTGTAAATAAGTTACACCAAGGAGGAGACACTTTAACAT	2768
QY	328	tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaTh	348
Db	2769	GCTGTCTTCTGACTTGGAGAGCTGAGAGAAAACCTTAGCAGATATCGAGGCAAAA-	2823
QY	348	rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaLeuValLysGlnI	368
Db	2824	-----TTTAGAGAGAA	2834
QY	368	eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr	388
Db	2835	AGATCAGAGAGAGAGAGCTGATAAGGCAAGGAAAAACTG-----	2877
QY	388	gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi	408
Db	2878	-----GAAAAATGAC-----ATTGCAGAAAATAATGAAG-----	2904
QY	408	sThrGlnGlyLeuGlnPheAspCysPheGlyThrAsnSerLysGlnSerValValSerTh	428
Db	2905	-----ATCTCAGGAGATAACTCTCTTCAG-----CTGAC	2933
QY	428	ryLysLeuValAsnGlu---SerLeuSerGlyAlaThrValArgAspGluLysIleAsn---	446
Db	2934	AAAAATGACGATGATTAACGCTCGAAGAGAGAGATGTAGAAGAAATACAGCTAAACT	2993
QY	447	-ThrLysSerMetArgAsnSerSerGluAspGluGluCysMetThrGluLysArgCysSe	466
Db	2994	TACAAAGGCTAATGAAAATGCAAGT-----	3018
QY	466	rHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTy	486


```
Db 3019 -----TTCTGCAAAAAGATTGAGCATGCTCTCAAGC 3056
Qy 486 rLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysMetGluLeuLeuAs 506
Db 3057 TGAACAGCAGCAGCAAGAGCAGCTAAAGAGCATGAGGAAGAAAGAAAGATTGGAGAG 3116
Qy 506 nMetTyrAlaAspLysGlnLysLeuGluThrSer-----LysSerValG1 522
Db 3117 GAAATTGTCGACCTCGAAAGAAAGAAATGGAACCAAGCCACACCAAGCTGTGAGGAGTGA 3176
Qy 522 uAlaAlaValAlaArgIleThrCysSerArgThrSerThrGlnValGlyAspLys---- 540
Db 3177 AGCCAGGTATGAGAGCCACT---TCTGAGACAAAACCAACATGAGAAATCTCTACA 3233
Qy 541 -----LysLeuLeuAspHisAsnTyrGluArgLysPheAspGluIleLysSe 556
Db 3234 GAACCTCCAGAGACGCTGCTGAC-----ACAGAGGACAAGCTGAAGGGCGCACGGGA 3287
Qy 556 rGluLysAsnGluCysLeuLysSerLeuGluGlnMetHisGluValAlaLysLysLe 576
Db 3288 GGAGAACAGTGGCTGCTGAGGAGCTGGAGGAGCTGAGAAAGCAAGCCGACAAAGCCAA 3347
Qy 576 uAlaGluAspGluAla 581
Db 3348 AGCTGCTCAACAGCG 3363

RESULT 15
US-10-831-704-79
; Sequence 79, Application US/10831704
; Publication No. US20050100931A1
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND
; TITLE OF INVENTION: ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/10/831,704
; CURRENT FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US/10/155,653
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US/09/220,132
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 5857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-831-704-79

Alignment Scores:
Pred. No.: 3,85e-05 Length: 5857
Score: 159.00 Matches: 136
Percent Similarity: 39.97% Conservative: 107
Best Local Similarity: 22.37% Mismatches: 211
Query Match: 4.94% Indels: 155
DB: 21 Gaps: 28

SEQ3_K705_COPY_650_1270 (1-621) x US-10-831-704-79 (1-5857)
Qy 31 GluArgThrLysIlePheArgLeuTyrSerValCysThrValGluGluLysAlaLeuIle 50
Db 1840 GAAGAAACATCATCAGAAAGAGATAAAGGCTCTGTATACCGCCACGAAAG-----1890
Qy 51 LeuAlaArgGlnAsnLysArgGlnAsnLysAlaValGluAsnLeuAsnArgSerLeuThr 70
Db 1891 CTTTCAAGAGAACACGATGATTGAAAGCAAGCAGCTGAGCATGCCCAAGAGAACTCA 1950
Qy 71 HisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPheAspLysLeuAspHisSer 90
```

```
Db 1951 GATGTGATAGCTCTATGGAAGTCC-----AAACTG-----1980
Qy 91 SerGluThrProAspSerGlyValSerPheGluGlnSerIleMetAspGlyValIleHis 110
Db 1981 -----GAGACTGCCATCGCATCCACAGCAGCGCATGGAAGAACTGAAGSTA 2028
Qy 111 GluPheSerSerIleLeuSerSerLysGlyGlyGluGluAsnGluValLysLeu-----128
Db 2029 TCTTTTCAGCAAAAGGGCTTGGAAACAGAGACGCGAGAAATTTGCTGAACTAAACACACAATA 2088
Qy 129 ---CysLeuLeuLeuAlaLysHisAlaGlnGlyThrTyrSerSerAspSerThrLeu 147
Db 2089 GAGAAATGAGACTAGATTACCAACACAGAAATAGAAAAATTTGCAGAAATCAACAAAGACTCT 2148
Qy 147 uPheGlyGluAspHisIleLysLeuSerAspGluGluSerProAsnIlePheTrpSer-- 166
Db 2149 GAACGGCTGCCCATCTTAAGAGA-----TGGAGGCC 2181
Qy 167 -----LysLeuLeuGlyGlyLysAsnProMetTrpLys-TyrProS 180
Db 2182 TTGAGGGCTAAACTGATGAAGTTATTAAAGAAAGAAAGAAACAGTCTGGAAGCCATCAGG 2241
Qy 180 erAspThrProGln-----ArgAsnArgLysArgValGlnTyrP 193
Db 2242 TCGAAACTGCAACAAAGCAGAACCCAGCATCTCGTAGAATGGAAGACACGTTTAAACAAA 2301
Qy 193 heGluGlySer-GluAlaSerProLysThrGlyAspGlyGlyAsnAlaLys-----Lys 210
Db 2302 TTACAGGAGCTGAATAAAGGTAAAGGAGCTAGAGGTACTGCAAGCCAAATCAATGAA 2361
Qy 211 ArgLysLysAlaSerAspValThrAspPro---ArgValThrAspProProValAsp 229
Db 2362 CAACCAAGGTTATTGATAATTTTACATCACAGCTCAAGCTACTGCAAGAAAGCTCTTG 2421
Qy 230 AspAspGlu-----ArgLysAlaSer-----GlyLysAspHisMetGlyAlaLeuGlu 245
Db 2422 GATCTTGATGCACTTCGGAAGCCAGTTCGGAAGGTAAATCGGAATGAAGAAACTTAGA 2481
Qy 246 SerPro-----LysValIleThrLeuGln-----Ser 254
Db 2482 CAGCAGCTTGAGCGCAGCTCGAAGAACAGATTAAACATTTAGAGATTGAAAGAAATGCTGAA 2541
Qy 255 SerCysLysSerSerGlyThrAspGlyThrLeuAspGlyAsnAspAlaPheGlyLeuTyr 274
Db 2542 AGTAGCAGGCTAGTAGCATTTACCAGAGAGCTCCAGGGGAGAGAG---CTAAAGCTTACT 2598
Qy 275 SerMetGlySerHisIleSerGlyIleProGluAspMetLeuAlaSerGlnAspTrpGly 294
Db 2599 AACCTTCAGGAAATTTGAGT-----GAAGT-CAGTCAAGTGAAGAGAGACTTTGGA 2648
Qy 295 LysIleProAspGluSerGlnArgArgLeu-----HisThrValLeu 308
Db 2649 AAAGAACTTCAGATTGTAAGAAAGAAAGTTTCTGTAAGCTTCAGAGGAGCGAGCTCTGT 2708
Qy 309 LysProLysMetAlaLysLeuCysGln-ValLeuHisLeuSerAspAlaCysThrSerMe 328
Db 2709 TCAGAGAGACTATGCAAGAACTGTAATAAGTTTACACCAAAAGGAGAGAAACAGTTTAACT 2768
Qy 328 tValGlyAsnPheLeuGluTyrValIleGluAsnHisArgIleTyrGluGluProAlaTh 348
Db 2769 GCTGTCTTCTGACTTGGAGAACTGAGAGAAACTTAGCAGATATGCGAGGCAAAA-----2823
Qy 348 rThrPheGlnAlaPheGlnIleAlaLeuSerTrpIleAlaAlaLeuLeuValLysGlnI 368
Db 2824 -----TTTAGAGAGAA 2834
Qy 368 eLeuSerHisLysGluSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerAr 388
Db 2835 ACATGAGAGAGAGAGAGAGCTGATTAAGGCAAGGAAAGAAACTG-----2877
Qy 388 gValGluValAspTyrIleTyrSerIleLeuSerCysMetLysSerLeuPheLeuGluHi 408
```

```
Db      2878 -----GAAATGAC-----ATTGCAGAAATAATGAAG----- 2904
QY      408 sThrGlnGlyLeuGlnPheAspCysPheGlyThrAenSerLysGlnSerValSerTh 428
Db      2905 -----ATGTCAGAGATAACTCTTCTCAG-----CTGAC 2933
QY      428 rLysLeuValAenGlu---SerLeuSerGlyAlaThrValArgAspGluLysIleAsn-- 446
Db      2934 AAAAAATGAACAGTGAATACGCTCTGAAGAGAGAGATGTAGAAGATTACAGCTAAACT 2993
QY      447 -ThrLysSerMetArgAenSerSerGluAspGluGluCysMetThrGluLysArgCysSe 466
Db      2994 TACAAAGGCTAATGAAATGCAAGT----- 3018
QY      466 rHisTyrSerThrAlaThrArgAspIleGluLysThrIleSerGlyIleLysLysLysTy 486
Db      3019 -----TTTCTGCAAAAAAGTATTGAGGACATGACTGTCAAAGC 3056
QY      486 rLysLysGlnValGlnLysLeuValGlnGluHisGluGluLysLysMetGluLeuLeuAs 506
Db      3057 TGAACAGAGCCAGCAAGAGCAGCTAAAAAGCATGAGGAAGAAAGAAAGAAATTGGAGAG 3116
QY      506 nMetTyrAlaAspLysLysGlnLysLeuGluThrSer-----LysSerValGl 522
Db      3117 GAAATTGTCGGACCTGGAAGAAAGAAATGGAACAGCCACCAACCAAGTGTCAAGAGCTGAA 3176
QY      522 uAlaAlaValIleArgIleThrCysSerArgThrSerThrGlnValGlyAspLeu---- 540
Db      3177 ACCCAGGTATGAGAGAGCCACT---TCTGAGACAAAAACCAAGCATGAAGAATCCTACA 3233
QY      541 -----LysLeuAspHisAenTyrGluArgLysPheAspGluIleLysSe 556
Db      3234 GAACCTCCAGAGAGCGTCTCTGGAC-----ACAGAGGACAAGCTGAAGGGCGCACGGGA 3287
QY      556 rGluLysAenGluCysLeuLysSerLeuGlnMetHisGluValAlaLysLysLysLe 576
Db      3288 GGAGAACAGTGGCTTGGCTGAGAGAGCTGGAGAGCTGAGAAAGCAAGCCGACAAAGCCAA 3347
QY      576 uAlaGluAspGluAla 581
Db      3348 AGCTGCTCAACACGCG 3363
```

Search completed: July 12, 2005, 00:57:15

Job time : 1049.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 23:52:39 ; Search time 14.6667 Seconds
(without alignments)
4073.905 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3213	99.9	2254	D86215	protein T6D22.14 [
2	163.5	5.1	1465	T23056	chromodomain helic
3	160	5.0	1738	C84507	hypothetical prote
4	152.5	4.7	1330	H89567	protein T08A9.1 [i
5	149.5	4.6	992	T46337	hypothetical prote
6	148	4.6	3187	JC5837	364K Golgi complex
7	147	4.6	962	C71617	SERA antigen/papai
8	147	4.6	3225	I52300	giantin - human
9	145	4.5	1130	T34081	hypothetical prote
10	144.5	4.5	1392	A43336	microtubule-veicli
11	143.5	4.5	1290	A55094	chromosomal protei
12	143	4.4	1269	F84730	probable myosin he
13	142.5	4.4	1427	S22695	restin - human
14	142.5	4.4	1940	A29320	myosin heavy chain
15	142	4.4	768	T02572	hypothetical prote
16	141.5	4.4	1927	A59236	embryonic muscle m
17	141.5	4.4	1999	S21801	myosin heavy chain
18	141.5	4.4	3147	T18674	hypothetical prote
19	141	4.4	741	S39082	myosin heavy chain
20	140.5	4.4	1961	A61231	myosin heavy chain
21	140.5	4.4	1972	A41604	myosin heavy chain
22	140	4.4	3259	A56539	giantin - human
23	139.5	4.3	1939	I48175	myosin heavy chain
24	139.5	4.3	2116	A26655	myosin heavy chain
25	139	4.3	853	T51505	hypothetical prote
26	139	4.3	1388	S74245	serine/threonine-s
27	138.5	4.3	1422	T18404	chromatin remodell
28	136.5	4.2	955	S24348	myosin heavy chain
29	136.5	4.2	1662	H71402	probable kinesin -

30	136	4.2	750	2	T38435	coiled coil protei
31	136	4.2	1039	2	S18199	myosin heavy chain
32	136	4.2	1938	2	A59293	skeletal myosin he
33	136	4.2	1939	2	T18372	repeat organellar
34	136	4.2	1957	2	T38077	hypothetical coile
35	135.5	4.2	1005	2	B87898	protein F10G8.8 [i
36	135.5	4.2	1041	2	T20730	hypothetical prote
37	135.5	4.2	1128	2	G86266	hypothetical prote
38	135.5	4.2	1997	2	F71607	DNA helicase II BR
39	135.5	4.2	2168	2	T30171	ninein - mouse
40	135	4.2	474	2	T50258	hypothetical coile
41	135	4.2	1186	2	C64588	cag pathogenicity
42	134.5	4.2	784	2	D96564	myosin-like protei
43	134.5	4.2	881	2	T52601	squamosa promoter
44	134	4.2	806	2	T08932	hypothetical prote
45	134	4.2	936	2	S39083	myosin heavy chain

ALIGNMENTS

RESULT 1

D86215
protein T6D22.14 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D86215
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.E.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D86215
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2254 <STO>
A;Cross-references: UNIPROT:Q9LN02; GB:AE005172; NID:98778840; PIDN:AAF79839.1; GSPDB:G
C;Genetics:
A;Gene: T6D22.14
A;Map position: 1

Query Match 99.9%; Score 3213; DB 2; Length 2254;
Best Local Similarity 99.8%; Pred. No. 6.8e-186;
Matches 620; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	RADAFILFGSSLNPSHDVXHVKIEKIESCSERTKIPRLYSVCTVEEKALILARQNKQNK 60
DB	883	RADAFILFGSSLNPSHDVXHVKIEKIESCSERTKIPRLYSVCTVEEKALILARQNKQNK 942
QY	61	AVENLNRSITALLMWSYLFDKLDHFSSTPDSGVSEFQSDGVIHFEFSSILSSKG 120
DB	943	AVENLNRSITALLMWSYLFDKLDHFSSTPDSGVSEFQSDGVIHFEFSSILSSKG 1002
QY	121	GEENEVKLCILLEAKHAQGYSSDSTLFGEDHKLSDSESPNIFWSKLLGGKQPMWKYP 190
DB	1003	GEENEVKLCILLEAKHAQGYSSDSTLFGEDHKLSDSESPNIFWSKLLGGKQPMWKYP 1062
QY	181	DTPQRNRKVVQVFESEASPKTDGGNAKKRKKASDDVTDPRVTDPPVDDDERKASGKH 240
DB	1063	DTPQRNRKVVQVFESEASPKTDGGNAKKRKKASDDVTDPRVTDPPVDDDERKASGKH 1122
QY	241	MGALESFKVITLQSSCKSGTDTLGDNDAGFLYSMGSHISGIPEDMLASQDWGKIPDES 300
DB	1123	MGALESFKVITLQSSCKSGTDTLGDNDAGFLYSMGSHISGIPEDMLASQDWGKIPDES 1182
QY	301	QORLHTVLPKPKAKLCQVHLSDACTSMYGNFLEYVIEHRIYEEPAITFOAFQIALSWI 360

```
Db      1183 QRRLLHTVLKPKMAKLCQVHLHSDACTSMVGNFLEYVVIENHRIYEBPATTFOAQIALSWI 1242
Qy      361 AALLVKQILSHKESVRANSELAFKSRVEVDYIYISLSCMKSLFLEHTQGLQFDCFCGTN 420
Db      1243 AALLVKQILSHKESVRANSELAFKSRVEVDYIYISLSCMKSLFLEHTQGLQFDCFCGTN 1302
Qy      421 SKQSVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTS 480
Db      1303 SKQSVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECWTEKRCSHYSTATRDIEKTS 1362
Qy      481 GIKKKYKKQVKLVQEHBEKKMELLNMYADKKQKLETSKVAAVIRITCSRTSTQVGD 540
Db      1363 GIKKKYKKQVKLVQEHBEKKMELLNMYADKKQKLETSKVAAVIRITCSRTSTQVGD 1422
Qy      541 KLLDHNRYERKFEIKSEKNECLSKLQMHVAVAKKLADEACWINRIKSWAAKLVKCVPI 600
Db      1423 KLLDHNRYERKFEIKSEKNECLSKLQMHVAVAKKLADEACWINRIKSWAAKLVKCVPI 1482
Qy      601 QSGNNKHFSGSSNISQAPDV 621
Db      1483 QSGNNKHFSGSSNISQAPDV 1503

RESULT 2
T23056
Chromodomain helicase H06001.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23056
R:Barlow, K.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19665
A:Accession: T23056
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1465 <WIL>
A:Cross-references: UNIPROT:O17909; EMBL:Z92970; PIDN:CAB07481.1; GSPDB:GN00019; CBSP.H0
A:Experimental source: clone H06001
C:Genetics:
A:Gene: CESP:H06001.2
A:Map position: 1
A:Introns: 4/1; 91/3; 170/3; 377/3; 494/3; 1046/2; 1099/3; 1242/3; 1298/1
C:Superfamily: Chromodomain helicase CHD1; chromobox homology

Query Match          5.1%; Score 163.5; DB 2; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.07;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

Qy      2 ADAFILGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVVEKALILARQNKQKA 61
Db      799 ADTVIIFDSWNPQNDLQMSRAHRIQTGTNVIYRLVTKGSVEE-----EI 845
Qy      62 VENLARSLLTHALLMWGASYLFDKLDH-----FHSETPDGSGVSEQSTMDGV 108
Db      846 VERARKKLV-----LDHLVTQRMDTGTVLSKNATAGSVPPDKQELSAI 891
Qy      109 IHEFSSI-----LSSKGGSENEV-----KCLCLLEAKHAQGYTSSDSTLFGF-DHIKLSDE 160
Db      892 L-KFGAVELFKEGESEQPEVDIDRLMGAETREAEEBVMKENELSSFKYANPAIDEE 950
Qy      161 PNIF-----WSKLLGKNPMWKYPSTDPQNRKRV-----Q 191
Db      951 KOIAATADEWAAII-----PBDNRNRIIEERMKELAEMLNAPRQKQPIQP 997
Qy      192 YPEGSEA-SPTKGDGNAKKRKKASDDVTPRV-----TDPVDDDERK 234
Db      998 VVEDDDGDDDEEDDTGKKKKKXKAVGNFTIPEIKRFIKRFRKPSMPLNLEEIAQAELE 1057
Qy      235 ASGKHMGALGSPKVITLQSSCKSGTGDGLDGNDAFLGYSMGSHISGIPEDMLASQDWG 294
Db      1058 EHSTDMMKKL-----VESLSEACKKADE--FDSNEKNG-----DAG 1092
Qy      295 KIPDESQ-----RRLHTVLKPKMAKLCQVHLHSDACTSMVGNFLEYV 336
```

RESULT 3

C84507

hypothetical protein At2g13370 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C:Accession: C84507

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84507

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1738 <STO>

A:Cross-references: UNIPROT:Q9SI41; GB:AE002093; NID:G4733988; PIDN:AAD28668.1; GSPDB:GN

C:Genetics:

A:Gene: At2g13370

A:Map position: 2

C:Superfamily: Chromodomain helicase CHD1; chromobox homology

Query Match 5.0%; Score 160; DB 2; Length 1738;

Best Local Similarity 17.4%; Pred. No. 0.14;

Matches 133; Conservative 102; Mismatches 242; Indels 286; Gaps 28;

Qy 2 ADAFILGSSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVVEKALILA-----52

Db 1041 ADTVIIFDSWNPQNDLQMSRAHRIQQGVVNIYRFTVTSKSVBEILERAERKMVL 1100

Qy 53 -----RQNKQKAVENLNRLTHALLMWGASYLFDKLDHFSHSETPDGSGVSE 103

Db 1101 VIQKLNABGRLEKRETKKGSNFDKNLSAILRFGABELFKE-----1141

Qy 104 IMDGVIHEFSSILSKGGEENEVKCLL-----LEAKHAQGYTSSDSTLFGEDH 152

Db 1142 -----DKNDEESKRLLSMDIDEILERAQVEEKY---TDETEHELLGAPK 1184

Qy 153 I-KLSDESPNIFWSKLLGKNPMWKYPSTDPQNRKRVQVPEGSEASPKTGDDGNNAKR 211

Db 1185 VANFCNAEDDGSFWSRWIKPDSVVTAEBALAPRAARNKTSYVDP SHPD-----RTSKRK 1238

Qy 212 KKASDDVTPRVTPDPPVDDDERKASGDHMGALGSPKVITLQSSCKSGTGDGLDGNDAF 271

Db 1239 KKGSE-----PPEHTERSQRKRTKTEYFVPTPLLEGTSAQVRGWSYGNLPRDQA 1288

Qy 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESQRRLHTVLKPKMAKLCOV 318

Db 1289 RPYRTVMKFGHNQMACIAEEVGGVVE-----AAPEEAQVELFDA-----1328

QY 319 LHLSDACTSV--GNF-----LEVYIENHRIYEEPATTFQ 351
Db 1329 --LIDGCKSVETGNFPGVLDFFGVPVKANELLKRVQGLQLLSKRISRYNDPISQFR 1386
QY 352 AFQALSWIA-----ALLVKQILSH--KESLVANSSEAF--KCSRVVE 391
Db 1387 ----VLSYLPDSNWSKCGNQDDARLLGLYHGFNGWEKIRLDSGLTKIAPVEL 1442
QY 392 DYIYSILSKMSLFLEHTQGLQDFP-----GTNSKQSVVSTKLVBNSLSG---ATVRDE 443
Db 1443 QHETFLPRAPNLKERATALLEMELAAAGKNTNAKASRKNKVKVNDLINFQKAPADR 1502
QY 444 K-----INTK-----SWNSSE-----DEE 458
Db 1503 RGKSGPANVLLSTKQGPRTQKAEPLVKEGEMSDGGEVYEFQEKQWMEWCEDVLADE 1562
QY 459 CMTKEKCHSVYATRD1--EKTISGKKYK--KOVOKLVQHEEKKMELLNMYADKKQ 513
Db 1563 IKTLGLRLOTTSDADLPKGVLPKIRRYLEILGRRIDAIVLEHEE-----DLYQDRM 1616
QY 514 KLETSKVEAAVIRITCSRTSTQVGDULKLDHNYERKFDRIKSEKNECLSKLEQMHEVAK 573
Db 1617 TMEL-----WNVVSTFSNLGDR-----LNQIYSK 1643
QY 574 KLADEACWINRIKSWAAKLKVCVPIQSGNNKHFSGSSNISQ 616
Db 1644 QKEEBEGV-----GPSHLNGSRNFOR 1665

RESULT 4
H89567
protein T08A9.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: H89567
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; PMID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: H89567
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1330 <STO>
A;Cross-references: UNIPROT:Q22342; GB:chr_X; PIDN:AAA81420.1; PID:g1065501; GSPDB:GN000
C;Genetics:
A;Gene: T08A9.1
A;Map position: X

Query Match 4.7%; Score 152.5; DB 2; Length 1330;
Best Local Similarity 19.0%; Pred. No. 0.28;
Matches 132; Conservative 112; Mismatches 225; Indels 227; Gaps 33;

QY 58 QNKAVENLNKSL---THALLMAGSYLPKLDHF-----HSETPDGSGVSPQSIM 105
Db 446 EKSREQLGRKLKHLFLHF---PGLFNLPEFYVYKAPLEKYDTDLNIAKEYIKDLR 501
QY 106 DGV-----IHFSSILSSKSGEENEVKCLLLEAKHAQGYSSDSTLFEGBHKL 154
Db 502 DALPELEPFLKVLTPNVASKLATGSKQH-----PSFGQTRVES 540
QY 155 -LSDEE-----SPNIFWSKLLGKNP-----MWKYPSTPQNRKR-----VQ 191
Db 541 FLTDEPWRMARSHFNYSIPAULSDGDSFPVQPLMCRSPENALRESQCRSPYVPSLQ 600
QY 192 YFGSEASPKTGDGNAKGRKASDDVDTPRVTDPVYD--DDEKASGKD--HMGAL--- 244
Db 601 QLEGLD-GPAPG-----SSAPISINSTSSRNPNKQSGRQSQSSODLHHVGSVSS 649
QY 245 -----ESP-----KVITLQSSCKSGTGTLDGNDAP-GLYSMSHISGIPED-MLAS 290

Db 650 DTSLLGHETPVKMEDTVMTENKSELNLSNMNDSIDSLDVEHYHIDYSAEDILLGL 709
QY 291 QDWGKIPDSQRRLHTV-----LKPMAKLCQVHLHS-----DACTSMVGNF 332
Db 710 SDHTVQLEDAVKLSKMQNIVANMTPRNEKLFYQLHQTAPRDPFIIITHDGDKLNRVKEL 769
QY 333 -----LEVYIENHRIYEEPATTFQAFQALSWIAALLVKQILSHKESLVANSSEAFK 386
Db 770 GESVESLELENKSTKTE-----SNLEAKLAEMTVSHKKEIENTQAE----- 811
QY 387 SRVEDVIYSILSKMSLFLEHTQGLQDFP-----GTNSKQSVVSTKLVBNSLSGATVRDESKIN 446
Db 812 -----CIRMSVEFE--LMTDSMSQSKQEIQESKOREIEELKAKLEKQNIH 856
QY 447 TKSMRNSDEECMTKEKCHSVYATRDIEKTIISGIIKKYKQVQKLVQ----- 495
Db 857 EKALRNDPYSEE-----YKRTLTAETR--AELEKEPKQRIEVTITKAHECKKDEAFA 905
QY 496 -----EHEEKOMELLNMYADKKQKLET--SKSVEAAVI-----RITCS 531
Db 906 ROEKTLEIENRVLSSENESSKSKLEAMNR--EKEQLESIRQMPEGEVILDFEFNALJES 962
QY 532 RTSTQVGD1-----KLLDHNYERKFDRI--KSEK-NECLSKLEOMHEVAKKK 575
Db 963 QPSAQIKDRFAIRSRMERSTQCKLDTNKVLLIDNLIKVKLQQLVNEREQYRDVIKSQ 1022
QY 576 LADEACWINRIKSWAAKLKVCVPIQSGNNKHFSGS 611
Db 1023 PGGDQV--LQFLENSTPTVPVDMEAVFNELTNFSGS 1056

RESULT 5
T46337
hypothetical protein DKFZp43402413.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46337
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23037
A;Accession: T46337
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-992 <AAA>
A;Cross-references: UNIPROT:Q9NTH6; EMBL:AL137265
A;Experimental source: adult testis; clone DKFZp43402413
C;Genetics:
A;Note: DKFZp43402413.1

Query Match 4.6%; Score 149.5; DB 2; Length 992;
Best Local Similarity 19.0%; Pred. No. 0.29;
Matches 125; Conservative 98; Mismatches 219; Indels 215; Gaps 26;

QY 122 ENEVKCLLLEAKHAQGYSSDSTLFGEDHKLSDSESPNIWFKLGG----- 171
Db 91 EKSEPKICIRNLVTPRADPT-----GSEPAKASEKAEPE---DTVDAGEEGSRREAA 139
QY 172 KNPWKY-----PSDTPQ-----RNRKRVQVPEGSEASPKTGDGNAKGRKASDDVD 220
Db 140 KEPKKAASALEGSSDASQELISEHMKEPQLSDSTASDPKSPHGLDFGRSRISEHLLD 199
QY 221 PRVTDP-----PVDDDERKASGKHOMGALES-----PKVITIQQSC 256
Db 200 VDLSPVLGACRAQAOPLGIEDKDSQSSQDELQSKGKLEERLSPPLPHEERAQSP 259
QY 257 KS-----SGTDGTLT-----GNDAFLYSM----- 276
Db 260 RSLATEEPPQGPQEPQEWKEABELGEDSAASLSLQLSLQREQAPSPPAACEKKGSHSQ 319
QY 277 -----GSHISGIPEDMLA----- 299
Db 320 ABELGFGQEAEDPEBKAVSPTPPVSPVRSSTEPVAPPEQLSEALKAMEEVAQVLEQ 379

Db 845 ELTGMRKLDMDTQQLNEETTAVVLFNDSIQEKIDKEKATINENWKLKSR--NEIAK 900

RESULT 10

A43336
microtubule-vesicle linker CLIP-170 - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C:Accession: A43336
R:Pierre, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.
Cell 70, 887-900, 1992
A:Title: CLIP-170 links endocytic vesicles to microtubules.
A:Reference number: A43336; MUID:92405160; PMID:1356075
A:Accession: A43336
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1392 <PI>
A:Cross-references: GB:M97501; NID:gl80621; PIDN:AAA35693.1; PID:gl80622

Query Match 4.5%; Score 144.5; DB 2; Length 1392;
Best Local Similarity 20.7%; Pred. No. 0.92;
Matches 124; Conservative 110; Mismatches 227; Indels 137; Gaps 28;

QY 31 ERTKIFRLSVCTVEKALILARQNRQKQKAVENLRSLTHALLMWGASYLFDKLDHFS 90
Db 535 EETHQKEIKALYTATEK---LSKENESLSKLEHANKENSVDIALWKS-----KL----- 581
QY 91 SETPDGVSFEQIMDGVIHFESSILSSKGGENEVK-----LCLLLEAKHAQGTYS 143
Db 582 ----ETAIASHQOAMEELKVSFKGLGTETAFAELKQTQIEKRLDYQHIEINLQOQDS 637
QY 144 DSTLFGED----HIKL-----SDEESPNI FWSKLLGGKPNMWPSPDTPQRNRK---RVQ 191
Db 638 ERAAHAKEMEARLAKMKVIEKENSLEAIRSKLDKAEQHLVEMEDTLNKLQEAIEKVK 697
QY 192 YFEGSEASPTGGNAKRRKASDDVTP-RVTDPPVDDDE--RKAS--GKHMGAL 246
Db 698 ELEVLQA-----KCNQTKVIDNFTSQLKATEBKLDLALRKASSEGKSEMKKLRQ 749
QY 247 P-----KVITLQ---SSCKSSGDTGLDGNDAFGLYSMSGSHISGIPEDMLASQDWCK 295
Db 750 QLEAAEKQIKHIEIKNAESSKASSITRELOREL-----KLTNLOENL---SEVSQ 798
QY 296 IPDESQRRLHTLKPMAKLCQVLHLSDACTSMVGNFLEYVJENHRIYEETPATTFAQFI 355
Db 799 VKETLEKEQL-ILKEKFAE-----ASEAVSVQSRMQETVKNLHQKEEQ-----FNM 844
QY 356 ALSWTAALLVKQILSHKESLVRANSELAFKCSRVEVDYIYSILSCMKSLFLEHTQGLQPD 415
Db 845 LSSDLEKL--RENLADEAKFREKDEREQLIKAKKEKLNDIAEIMK----- 889
QY 416 CFTGNSKQSVSTKLWNE-SLSGATVRDEKIN-TKSMRNSDEECWTEKRCSHYSTA 473
Db 890 MSGDNSSQ---ITQWDELRLKRDVVEQLKLTKANENAS----- 927
QY 474 DIEKTISGIKKYKQVQLVQHEHEKQMLLMYADKKQKLETS-----KSVEAAVIRIT 529
Db 928 FLQKSIEDMTVKAESQSQAAKHEEKEKLEKRLSDLEKQMETSHNOQELKARYERAT 987
QY 530 CSRTSTQVGD-----KLLDHNRYKFPDIKSEKNECLLSQEMHEVAKKLADEA 581
Db 988 -SETTKTHEILQNLQKLTLLD--TEDKLKGAREENSGLLQLEFLRQKAEKAAQTA 1042

RESULT 11

A55094
chromosomal protein XCAP-C - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 09-Jul-2004
C:Accession: A55094
R:Hirano, T.; Mitchison, T.J.
Cell 79, 449-458, 1994
A:Title: A heterodimeric coiled-coil protein required for mitotic chromosome condensatio

A:Reference number: A55094; MUID:95042742; PMID:7954811

A:Accession: A55094

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1290 <HIR>

A:Cross-references: UNIPROT:P50532; GB:U13673; NID:9563811; PIDN:AAA64679.1; PID:9563812

C:Superfamily: chromosome segregation protein SMC1

C:Keywords: chromosomal protein; DNA condensation; heterodimer

Query Match 4.5%; Score 143.5; DB 2; Length 1290;
Best Local Similarity 18.2%; Pred. No. 0.95;
Matches 125; Conservative 105; Mismatches 227; Indels 231; Gaps 28;

QY 16 HDVGHVERKIKIESCSERTKIFRLSVCTVEEKAL-----ILARQNRQKQKAVENLR 67
Db 400 HSKSVKVKLQKLOKQDKKEVDLKNVPANSQKIIAETNKKDLLEKQKEBEKLKNVMD 459
QY 68 SLTHALLMWGASYLFDKLDHFSSETPDGVSFEQIMDGVIHFESSILSSKGGENEVK 127
Db 460 SL-----KKET--QGLQEEKVEKEKELMEISKTVN----- 487
QY 128 LCLLLEAKHAQGTYSSTLFGEDHI KLSDESPNIFWSKLLGGKPNMWPSPDTPQRNR 187
Db 488 ----EARKMDVAQSELDIYLSRH-----NSALSQNLKAKEAL-NTASATLKERR 532
QY 188 KRVQYFEGSEASPTGGNAKRRKASDDVTDTP----- 221
Db 533 AAIKELE--TKLPK--DEGLKREKELESIVSEBGNIKNOVRELQKVVEARSSILSANR 588
QY 222 ---RVTDPPVDDDERKASGK-----DHMGALSPKVITLQSSC----- 256
Db 589 SRGKVLDAI---QOKSGKIPGIFRGLDGLGAIDKDYDAISSCGALDHIWVDITDPA 645
QY 257 -----KSSGTDGTLDGNDAFGLYSMSGSHISGIPEDMLASQDWGKIPDESQRRLHTVL 308
Db 646 QECVNFLLKQNVGVATFGLDKMKWKEGLNKIQTPENI PRLPDMVVKVQDEQ-----I 698
QY 309 KPMAKLCQVLHLSDACTSMVGNFLEYVJENHRIYEETPATTFAQFI ALSWTAALLVKQI 368
Db 699 KPAF-----YFALRD---TIVANNLD-----QATRVAFQKQKRWVVTLQGGI 738
QY 369 LSHK-----ESLVRANSELAPFKCSR----- 391
Db 739 IEQSGTMTGGGKVMKGMGSMVMVEISDDQLQKVENKLTDTTTRATEIQDKRAHLEEV 798
QY 392 DYIYSILSCMKSLFLEHTQGLQDFCGFNYSKQSVSTKLWNEISLGATVRDEKINTKSMR 451
Db 799 AKLRQATREMKNVTFEYKTASLQ-----SLSEQEVHLK-----AQVKELEVNVA 844
QY 452 NSSEDEECWTEKRCSHYSTATRDIEKTI SGIKKKYKQVQLVQ-EHEBEKQMLL-----N 506
Db 845 PDKNQ-----KQMEKNLETLKKEYKVAEKAGKAGVEAEVRLHKLIVDIN 889
QY 507 MYADKKQ--KLE-TSKSVSEAAVIRITCSRTSTQVGD LKLLDHNRYKFPDIKSEKNECL- 562
Db 890 NHLKRAQDQKLDVKTKIDECASAITKAQVS-----IKTADRNLKKSBEAVARTEKEIVA 944
QY 563 --KSLEOMHEVAKKLADEACWINRIK 588
Db 945 NDKSIEELTE-DLKKLEKATTVMNECK 971

RESULT 12

F84730
probable myosin heavy chain [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84730
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
eaus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: F84730

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1269 <STO>

A;Cross-references: GB:AE02093; NID:g6598483; PIDN:AAC6932.2; GSPDB:GN00139

C;Genetics:

A;Gene: At2g32240

A;Map position: 2

Query Match 4.4%; Score 143; DB 2; Length 1269;

Best Local Similarity 21.9%; Pred. No. 1;

Matches 110; Conservative 81; Mismatches 192; Indels 120; Gaps 20;

QY 150 EDHKLSDERSPNIFWSKLGGKPNMWPSPDTPQRRNRKVQVPEGSEASPKTGDGNAK 209

DB 2 EATQVTSSEVP-----VVKGRQRCKYCLDLLQAVNGEVPKKEEEE-----DGEFIK 50

QY 210 KRKASDDVTDPRTVDPVDDDERKASGKDHMGALSPKVTITLQSSCKSGTGTGLDND 269

DB 51 VEKEAFD-----AKDDAEKA---DHVPVEEQKEVI-----BRSSSGSQRELHESQ 92

QY 270 -----AFGLYSNGSHSIGIPEDMLASODWGKIPDRSQRRLHTVLKPKMAKLC 316

DB 93 EKAKELELERVAGELKRYESENTHLKDELLSAKELKETEKKHGDLVVQKQKEKIV 152

QY 317 QVLHLSDACTSMVGNFLEYVIENHRIYEEPT-----TFQAFQIALSWIAALLV----- 365

DB 153 E-----GEERHSSQLKSLDALQSHDAKDTELTEKAFDALGIELESSRKKLIELEGLK 208

QY 366 -----KQLSHKESLVRANSELAPKCSRVVDYIYSILSCWKSIL--FLEHTQGLQFDC 416

DB 209 RSAAEAQKPELHKQASHADSE-----SQKALEFSELLKSTKESAKEMERKASLQOEI 263

QY 417 FGNSKOS---VYSTKLVNESLSGATVRDEKINTKSMRNSSE-----D 456

DB 264 KELNEKSENEKVEALKSAGELAAVQEEALSKSLLETEKQVSTEMALIDELTQELE 323

QY 457 EECWTEKCRSHYSTATRDIEKTISGKKYKQV---QKLVQHEBKMBELLMYADKKQ 513

DB 324 QKASERPFKELSVLQDLDAQTKGLQAKLSEQEGINSKLAELKEKEL--LESLSKDQEE 382

QY 514 KLETS-----KSVEAAVIRITCS--RTSTQVGD-----LKLDHNYERKFDRI-- 554

DB 383 KLTANKEKLAELVLEKEALEANVAETSVNATVTEVCNELEBKLTSDENFS--KTDALLS 441

QY 555 -----KSEKNECKLSLEOMHEVA 572

DB 442 QALSNNSSELEKQKUSLEELHSEA 464

RESULT 13

S22695

restin - human

C;Species: Homo sapiens (man)

C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004

C;Accession: S22695; S19853

R;Billbe, G.; Delabie, J.; Bruggen, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;

EMBO J. 11, 2103-2113, 1992

A;Title: Restin: a novel intermediate filament-associated protein highly expressed in th

A;Reference number: S22695; MUID:92289675; PMID:1600942

A;Accession: S22695

A;Molecule type: mRNA

A;Residues: 1-1427 <BIL>

A;Cross-references: UNIPROT:P30622; EMBL:X64838; NID:g35998; PIDN:CAA46050.1; PID:g35998

C;Keywords: cytoskeleton

Query Match 4.4%; Score 142.5; DB 2; Length 1427;

Best Local Similarity 20.7%; Pred. No. 1.2;

Matches 124; Conservative 109; Mismatches 228; Indels 137; Gaps 28;

QY 31 ERTKIFPLVSVCTVEKALILARONKRONKAVENLARSLLTHALLMGASVLPFKLDHFS 90

DB 31 ERTKIFPLVSVCTVEKALILARONKRONKAVENLARSLLTHALLMGASVLPFKLDHFS 90

DB 570 ETHQKEIKALYTATEK---LSKENSLSKLEHANKENSVDIALMKS-----KL----- 616

QY 91 SETPDSGVSPFEQIMDGVIIHEFSSILSSKGEENEYK-----LCLLLEAKHAQGYSS 143

DB 617 ----ETAIASHQAMEELKVSFGKLGITETABPAELKTIQIEKWLVDYQHEIENLQNOQDS 672

QY 144 DSTLFGED----HIKL-----SDESPNIFWSKLGGKPNMWPSPDTPQRRNRK---RVQ 191

DB 673 ERAAHAKEMEARAKLMKVIKEKENSLEAIRSKLDAEQHLVEMEDTLNKLQEAETIKVK 732

QY 192 YPGSEASPKTGGGNAKKKASDDVTDP--RVTDPPVDDDE--RKAS--GKDHMGAL 246

DB 733 ELEVLQA-----KCNQOTKVIDNFTSQLKATEEKLDDALRKASSEKSEMKKLRQ 784

QY 247 P-----KVITLQ---SSCKSGTGTGLDNGDAFGLYSNGSHSIGIPEDMLASODWGK 295

DB 785 QLEAAEQIKHLIEKNAESSKASSITRELQGREL-----KLTNLQENL---SEVSQ 833

QY 296 IPDESORRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRIYEEPTTFQAFQI 355

DB 834 VKETLEKELQ--ILKEKFAE-----ASEEAVSVQRSMQETVNLKHQKEEQ-----FNM 879

QY 356 ALSWIAALLVKQILSHKESLVRANSELAPKCSRVVDYIYSILSCWKSILFLEHTQGLQFD 415

DB 880 LSSDLSEK---RENLADMEAKFREKDEREROLIKAKKLENDIAEIMK----- 924

QY 416 CFTGNSKQSVVSTKLVNE--SLSGATVRDEKIN--TKSMRNSSEDEECWTEKCRSHYSTATR 473

DB 925 MSGDNSSQ---LTKMDELRLKRDVEELQKLITKANENAS----- 962

QY 474 DIEKTSISGKKYKQVQKLVQHEBKMBELLMYADKKQKLETS-----KSVEAAVIRIT 529

DB 963 FLQKSTEDMTVRAEQSQEAAKHEKKELEKRLSDLEKQMETSHNQCOELKARYERAT 1022

QY 530 CSRTSTQVGD-----KLLDHNYERKFDRIEKEKNECKLSLEOMHEVAKKLAEDBA 581

DB 1023 -SETKTKEIILQNLQKLTLD--TEDKLGARENSGLQLEBELRQADKAKAAQTA 1077

RESULT 14

A29320

myosin heavy chain, fast skeletal muscle, embryonic [similarity] - chicken

N;Alternate names: myosin heavy chain, EFWI

C;Species: Gallus gallus (chicken)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004

C;Accession: A29320; B24124; A02990

R;Molina, M.I.; Kropp, K.E.; Gulick, J.; Robbins, J.

J. Biol. Chem. 262, 6478-6488, 1987

A;Title: The sequence of an embryonic myosin heavy chain gene and isolation of its cor

A;Reference number: A29320; MUID:87194881; PMID:3571266

A;Accession: A29320

A;Molecule type: DNA; mRNA

A;Residues: 1-1940 <MOL>

A;Cross-references: UNIPROT:P02565

A;Note: the sequence translated from the mRNA differs from that of the DNA in having 37

R;Gulick, J.; Kropp, K.; Robbins, J.

J. Biol. Chem. 260, 14513-14520, 1985

A;Title: The structure of two fast-white myosin heavy chain promoters. A comparative st

A;Reference number: A92507; MUID:86033956; PMID:2997212

A;Accession: B24124

A;Molecule type: DNA

A;Residues: 1-168 <GUL>

R;Kavinsky, C.J.; Umeda, P.K.; Sinha, A.M.; Elzinga, M.; Tong, S.W.; Zak, R.; Jakovic,

J. Biol. Chem. 258, 5196-5205, 1983

A;Title: Cloned mRNA sequences for two types of embryonic myosin heavy chains from chic

A;Reference number: A02990; MUID:83161144; PMID:6833296

A;Accession: A02990

A;Molecule type: mRNA

A;Residues: 1502-1546, 'A', 1548-1912, 'G', 1914, 'T', 1916-1940 <KAV>

A;Cross-references: GB:V00430; GB:J00892; NID:g63619; PIDN:CAA23712.1; PID:g808972

A;Experimental source: clone 251

C;Comment: The entire sequence of this fragment corresponds to the light meromyosin por

ides (heptads I-IV), that are characteristic of coiled coils.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 11, 2005, 10:23:56 ; Search time 5377.33 Seconds
(without alignments)
4395.843 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270

Perfect score: 3216
Sequence: 1 RADAFILFGSSLNPSHDVXH.....SGNNKHFGSSNISQNPADV 621

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/ROBINSON929/runat_11072005.102152.29186/app_query.fasta_1.2325
-DB=EST -QFMT=FASTAP -SUFFIX=rst -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=DISSIM2 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=ROBINSON929 @CGN 1.1.8678 @runat_11072005.102152.29186 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	413	12.8	654	8	BH989471 oem85a08
2	323.5	10.1	775	9	CL818501 OR_CBA003
3	313.5	9.7	764	9	CL795819 OR_CBA000
4	299.5	9.3	369	8	BH419713 BOG0869TF
5	298.5	9.3	798	9	CG167203 PUHJ72TD
6	228	7.1	525	4	BG833433
7	224.5	7.0	569	2	BE203168
8	221	6.9	482	4	BG833434
9	213.5	6.6	246	9	CC794080 SALK_0411

10	211.5	6.6	516	1	AU288479
11	210	6.5	668	5	BP133890
12	208.5	6.5	486	1	AI947519
13	201	6.2	799	8	B21272
14	198	6.2	560	8	BH788688
15	186	5.8	669	6	CA215647
16	183	5.7	942	8	B08967
17	181	5.6	529	5	BQ873535
18	180	5.6	969	9	CG396389
19	178.5	5.6	519	1	AU302228
20	168	5.2	690	8	BH978547
21	143.5	4.5	928	8	B11993
22	143.5	4.5	2865	3	AK014630
23	142.5	4.4	2079	3	AK017709
24	141.5	4.4	3927	9	CL971593
25	140.5	4.4	341	6	CB084203
26	140	4.4	2176	3	AK016726
27	139.5	4.3	1338	7	CO635655
28	138	4.3	2019	9	CL958706
29	138	4.3	5042	3	HS805132
30	137.5	4.3	3738	9	CL972961
31	135	4.2	4137	9	AY420057
32	135	4.2	7050	9	AY420513
33	134.5	4.2	848	8	BZ998864
34	134.5	4.2	4095	9	CL963545
35	134	4.2	1512	9	CL959694
36	133	4.1	2867	3	AK028516
37	132.5	4.1	1105	7	CR754231
38	132.5	4.1	6795	6	CL969745
39	132	4.1	668	6	CB526089
40	132	4.1	778	6	CA315071
41	132	4.1	2615	3	AK044987
42	132	4.1	3887	3	AK081567
43	131.5	4.1	3180	9	CL961241
44	131.5	4.1	3344	3	BC028681
45	131	4.1	2133	3	AK014977

ALIGNMENTS

RESULT 1

BH989471

LOCUS

DEFINITION

oem85a08.b1 B.oleracea002 Brassica oleracea genomic, genomic survey

654 bp DNA linear GSS 07-OCT-2002

sequence.

ACCESSION BH989471

VERSION BH989471.1

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE Delhaudy, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,

AUTHORS Nash, W., Rabinowicz, P.D. and Wilson, R.K.

TITLE Whole genome shotgun reads from Brassica oleracea

JOURNAL Unpublished (2002)

COMMENT Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Plate: oem85 row: a column: 08

Seq primer: -21UpOT forward

Class: shotgun

High quality sequence start: 20

High quality sequence stop: 551.

Location/Qualifiers

1. .654

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone_lib="B.oleracea002"

source

/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassaica oleracea TO1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Alignment Scores:
 Pred. No.: 1.69e-34 Length: 654
 Score: 413.00 Matches: 90
 Percent Similarity: 67.95% Conservative: 16
 Best Local Similarity: 57.69% Mismatches: 28
 Query Match: 12.84% Indels: 22
 DB: 8 Gaps: 2

SEQ3_K705_COPY_650_1270 (1-621) x BH9889471 (1-654)

QY 358 SerTrpIleAlaLeuLeuValLysGlnIleLeuSerHisLysGluSerLeuValArg 377
 |||||
 Db 253 AGTTGGATTGACGCTCGTAAAGCAACGTTAACCGCGAAGAACTTTTGGTCCGC 312
 |||||
 QY 378 AlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIle 397
 |||||
 Db 313 GC AAAATCGGAATTAGGTTCTGCTGAGAGAGAGGTAGATTATATATCTCTCT 372
 |||||
 QY 398 LeuSerCysMetLysSerLeuPheLeuGluHisThrGlnGlyLeuGlnPheAspCysPhe 417
 |||||
 Db 373 TTGTGCTGATGAGAGTATATCTCGGGCGCACACGAAGTTTGCGGTGATTAATG 432
 |||||
 QY 418 GlyThrAsnSerLysGlnSerValValSerThrLysLeuValAsnGlnSerLeuSerGly 437
 |||||
 Db 433 GGTACTAATCTTAAGCAACA-----AAGTGAAATGAAAGTCTCTCAGGG 477
 |||||
 QY 438 AlaThrValArgAspGlnLysIleAsnThrLysSerMetArgAsnSerSerGluAspGlu 457
 |||||
 Db 478 GATACAGTCGTCAGGAACAGAGTAACACAAAATCGATCGCT-----519
 |||||
 QY 458 GluCysMetThrGluLysArgCysSerHisTyrSerThrAlaThrArgAspIleGluLys 477
 |||||
 Db 520 -----AGCTCAGTAACAAAAGATGTCGAAAG 546
 |||||
 QY 478 ThrIleSerGlyIleLysLysLysTyrLysGlnValGlnLysLeuValGlnGluHis 497
 |||||
 Db 547 ACTATTAGCGCATCAAAATAAATGCAGTAAGAACTGCAGAGCTTGTAATAATCCAC 606
 |||||
 QY 498 GluGluLysLysMetGluLeuLeuAsnMetTyrAlaAspLysLysGln 513
 |||||
 Db 607 GAGGAAAAAAGGTGGATCTGATGATAGGAGTGTGTCGACAAAGACAG 654
 |||||

RESULT 2

CL818501
 LOCUS OR_Cba0036H03.r OR_Cba Oryza rufipogon genomic clone OR_Cba0036H03
 DEFINITION 3', genomic survey sequence.

ACCESSION CL818501

VERSION CL818501.1 GI:51064111

KEYWORDS GSS.

SOURCE Oryza rufipogon

ORGANISM

Oryza rufipogon
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 775)

AUTHORS Kim.H., Yu.Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
 Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
 Wing,R.

TITLE OMAP project

JOURNAL Unpublished (2004)

CONTACT: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA
 Tel: 520 626 9595
 Fax: 520 621 1259
 Email: http://genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0036 row: H column: 03

Seq primer: CAC TCA TTA GGC ACC CCA

Class: BAC ends.

FEATURES Location/Qualifiers

1..775
 /organism="Oryza rufipogon"
 /mol_type="genomic DNA"
 /db_xref="taxon:4529"
 /clone="OR_Cba0036H03"
 /tissue_type="young leaves"
 /dev_stage="2 week old seedlings"
 /lab_host="DH10B T1 phage resistant"
 /clone_lib="OR_Cba"
 /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
 drk treated 36 hrs before harvest"

ORIGIN

Alignment Scores:
 Pred. No.: 1.64e-24 Length: 775
 Score: 323.50 Matches: 85
 Percent Similarity: 50.79% Conservative: 43
 Best Local Similarity: 33.73% Mismatches: 105
 Query Match: 10.06% Indels: 19
 DB: 9 Gaps: 2

SEQ3_K705_COPY_650_1270 (1-621) x CL818501 (1-775)

QY 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
 |||||
 Db 6 GATGCCATCATATATGGTTCTGACAAACCCACTTAACGACTTGAAGCTCTTCAA 65
 |||||
 QY 23 LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
 |||||
 Db 66 AAGATCAAGATAGACTCAAAATTTGAGCGGTGAGCAATTTTCGCTTGTAACACTCCTTTT 125
 |||||
 QY 43 ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
 |||||
 Db 126 ACAGTTTGAGAGAGAGACTTTGTTCTTCGAGCGAGGCGCATTTGTTATTGATTAACACATA 185
 |||||
 QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTyrGlyAlaSerTyrLeuPhe 82
 |||||
 Db 186 CAAGATCTAAGGACTAGTTTAAACATTTCTTACTCAGATGGGGTGTGTCATTTCTCTTC 245
 |||||
 QY 83 AspLysLeuAspHisPheHisSerSerGluThrProAspSerGlyValSerPheGluGln 102
 |||||
 Db 246 AGTAGACTTGATGAGGTCCAGCAAGATGACCATGCAAGTAAAGTTCTGAAATGAGAGA 305
 |||||
 QY 103 SerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLysGlyGlu 122
 |||||
 Db 306 CACTTTATTGATGAAGTAATCGTTGAGTTCTTAAACAAAGTTTATCCACAACTGTTGAAGAC 365
 |||||
 QY 123 GluAsnGluValLysLeuCysLeuLeuGluAlaLysHisAlaGlnGlyThrTyrSer 142
 |||||
 Db 366 AGCAGTGAAGTACACAGGAAGTCCATAAGCAAGTAATATGATGGGGAACTATATCA 425
 |||||
 QY 143 SerAspSerThrLeuPheGlyGluAspHisIleLysLeuSerAspGlyGluSerProAsn 162
 |||||
 Db 426 AGAAATATTACTCTCTGATGGTGAAAGGAGGAATCTCTGTCTTGAAGGATAATCCAGCT 485
 |||||
 QY 163 IlePheTrpSerLysLeuLeuGlyCysLysAsnProMetTrpLysTyrProSerAspThr 182
 |||||
 Db 486 GAGTTCTGGTTAAATTTATTGGATGGAAGATCCCTCATGTGAGCTGTATATCTGAGCCA 545
 |||||
 QY 183 ProGlnArgAsnArgLysArgValGlnTyrPheGluGlySerGluAlaSerProLysThr 202
 |||||
 Db 546 CTGCAATCAAGAGTTACAAGTCCAGCATGGATGAGTCAATGCTCCGCTCGAGAA 605
 |||||


```

mRNA sequence.
ACCESSION EG833433
VERSION BG833433.1 GI:14191203
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 525)
Zealote, V.
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 951008 row: C column: 07.
FEATURES
Location/Qualifiers
1..525
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth
phases"
/lab_host="DH10B"
/clone_lib="951 - BMS tissue from Walbot Lab (GR)"
/notes="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
library was prepared by George Rudenko using poly (A)
selected RNA and Universal Riboclone cDNA Synthesis System
(from Megal). cDNA was synthesized using both random and
oligo(dT) primers in separate reactions and equipped with
EcoRI adaptors. Library was size-fractionated on agarose
gels (for insert size >400bp) and non-directionally cloned
into EcoRI-digested pUC19 vector. Blue/white selection on
carbenicillin-containing plates was used to recover
positive clones."
ORIGIN
source
Alignment Scores:
Pred. No.: 2 93e-14 Length: 525
Score: 228.00 Matches: 53
Percent Similarity: 63.87% Conservative: 23
Best Local Similarity: 44.54% Mismatches: 37
Query Match: 7.09% Indels: 6
DB: 4 Gaps: 2
SEQ3_K705_COPY_650_1270 (1-621) x BG833433 (1-525)
QY 3 AspAlaPheLeuLeuPheGlySerLeuAsnProSerHisAspVallyHisValGlu 22
Db 94 GATGCCATCATCATATTAGTGAATCCCATGACTTGGAGAGCCCTTCAG 153
QY 23 LysAlaLeuLeuSerGlySerGluArgThrLysAlaLeuPheArgLeuTyrSerValCys 42
Db 154 AGGGTCAGCATGAGTGCACATGCTGAACCTGTGCTGCTATTTTTCGTTTGTACTGCTCTTT 213
QY 43 ThrValGluGluLysAlaLeuLeuLeuAlaArgGlnAsnLysArgGlnAsnLysAlaVal 62
Db 214 ACAGTGAGGAAAGAGCTTATCTTCGAAAGCATGACCATATTTTGGATAGTATATAT 273
QY 63 GluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMetTrpGlyAlaSerTyrLeuPhe 82
Db 274 GTGAATATACCCCTAGCCATTTGCTCTCAGTTGGGGCGCATCATTTCTTCTTC 333
QY 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100
Db 83 AspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerPhe 100

```

```

334 AATAGACTTGAGGAGTTGCAGCAACATAGCTAC-----TCAATGTCTCTGGT 381
QY 101 GluGlnSerIleMetAspGlyValIleHisGluPheSerSerIleLeuSerSerLys 119
Db 382 GATGAGCTTTTCATGATAATGTCGATTTGGAGTCTTTTGACAAATTTACTCAGCAAG 438
RESULT 7
BE203168 569 bp mRNA linear EST 07-SEP-2000
LOCUS EST403190 KV1 Medicago truncatula cDNA clone pKV1-4N3, mRNA
DEFINITION sequence.
ACCESSION BE203168
VERSION BE203168.1 GI:8746439
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 569)
VandenBosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
Fraser, C.M.
ESTs from roots of Medicago truncatula 24 hours after inoculation
with Sinorhizobium meliloti
Unpublished (1999)
Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@chs.umn.edu
Texas A&M University, T262940e
TIGR sequence name: M1AK74TK
More information is available at:
http://chrysis.tamu.edu/medicago
Seq primer: SKnod (CTA gaa cta gta gat CC).
FEATURES
Location/Qualifiers
1..569
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone_lib="pKV1-4N3"
/tissue_type="Seedling roots"
/dev_stage="24 hours post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XL0LR"
/clone_lib="KV1"
/notes="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
ORIGIN
Alignment Scores:
Pred. No.: 8 12e-14 Length: 569
Score: 224.50 Matches: 64
Percent Similarity: 42.27% Conservative: 29
Best Local Similarity: 29.09% Mismatches: 78
Query Match: 6.98% Indels: 49
DB: 2 Gaps: 6
SEQ3_K705_COPY_650_1270 (1-621) x BE203168 (1-569)
QY 206 GlyAsnAlaLysLysArgLysLysAlaSerAspAspValThrAspProArgValThrAsp 225
Db 40 GGCACACAAATAAGCGCTTGAAGTAGTAAACAAT----- 75

```


Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

Email: eckersgalk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated intron of A1198060. Class: TDNA tagged.

cross. data suggest.

FEATURES . Location/Qualifiers
source 1. .246

```
1. 246
/organism="Arabidopsis thaliana"
/mol type="genomic DNA"
/ecotype="Col-0"
/db xref="taxon:3702"
```

`/clone="SALK_041106_30_85_x"
 /clone.lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines"
 /each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used
 can be found at http://signal.salk.edu/tDNA_protocols.htm`

ORIGIN

Alignment Scores:		
Pred. No.:	3.37e-13	246
Score:	23.50	50
Percent Similarity:	70.83%	1
Best Local Similarity:	69.44%	Conservative: 5
Query Match:	6.64%	Mismatches: 16
DB:	9	Indels: 2
		Gaps: 2

SEQ3 K705 COPY 650 1270 (1-621) x CC794080 (1-246)

QY 334 GluTyrValIleGluAsn-HisArgIleTyrGluProAlaThrPheGlnAlaPh 353
|||||:::|||||
Db 77 GAATACGTTGTTAATAATGCAT-----TT 100

Qy 353 eGinIleAlaLeuSerTriPileAlaAlaLeuValLysGlnIleLeuSerHisLysG1 373
|||
Db 101 TCAG-----AGATGATTCGACCTTTGGAAAAGCAAAATCTTACCCACAAAGA 151

373 uSerLeuValArgAlaAsnSerGluLeuAlaPheLysCysSerArgValGluValAspTy 393
152 ATCTCTGGTCCGTCGCAAAATCTGAATTAGCTTTCAAATGCTCTAGAGAAGAGGTGGAATTA 211

Qy 393 rIleTyrSerIleLeuSerCysMetLysSerLeu 404
Db 212 TATTTATTGGATATTGCTCGTCGAACAGTCTG 245	

RESULT 10					
AU288479					
LOCUS	AU288479	516 bp	mRNA	linear	EST 04-DEC-2002
DEFINITION	AU288479 zinnia cultured mesophyll cell equalized cDNA zinnia elegans cDNA clone Z2304, mRNA sequence.				

ACCESSION AU288479
VERSION AU288479.1 GI:24248599
KEYWORDS EST.

KEYWORDS	SOURCE	EST.
	Zinnia elegans	
	Zinnia elegans	
ORGANISM		

ORGANISM	REFERENCE
<i>Zinnia elegans</i>	1 (bases 1 to 516)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asteride; campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Zinnia.	

REFERENCE
1 (bases 1 to 516)
Demura, T., Tashiro, G., Horiguchi, G., Kishimoto, N., Kubo, M.,
Horiuchi, T., Shimizu, T., and Shimizu, T. (1990) *Journal of*

Demura, I., Iasuiro, G., Horiguchi, S., Kishimoto, N., Kudo, M.,
Matsuoka, N., Minami, A., Nagata-Hiwatashi, M., Nakamura, K.,
Okamura, Y., Sassa, N., Suzuki, S., Yazaki, J., Kikuchi, S. and
Fukuda, H.

TITLE Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT
Contact: Taku Demura

Morphogenesis Research Group
RIKEN Plant Science Center
1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9605
Fax: 81-45-503-9573

Email: demura@postman.riken.go.jp
This clone was obtained at our laboratory.
Seq primer: M13 forward

FEATURES
source

```

/organism="Zinnia elegans"
/mol_type="mRNA"
/cultivar="Canary bird"
/db_xref="taxon:34245"
/clone="22904"

```

```
/clone=Z2504
/tissue type="mesophyll cell"
```

ORIGIN

ORIGIN

Alignment Scores:	1.89e-12	Length:	516
Pred. No.:	211.50	Matches:	51
Score:	58.33%	Conservative:	46
Percent Similarity:	58.33%	Mismatches:	19
Best Local Similarity:	42.50%	Indels:	4
Query Match:	6.58%	Gaps:	1
DB:	1		

SEQ3 K705 COPY 650 1270 (1-621) x AU288479 (1-516)

Qy	299	GluSerGlnArgArgLeuHisThrValLeuLysProLysMetAlaLysLeuCysGlnVal	318
Db	82	GAAGGGGAATTGAACTTTAAGGACCTCTCGAAGCTTAAATACCTAACTATGTGAAGTC	141

Qy	319	L <u>e</u> uH <u>i</u> s <u>e</u> rS <u>e</u> rA <u>s</u> pA <u>a</u> C <u>y</u> s <u>t</u> rS <u>e</u> rM <u>e</u> tV <u>a</u> I <u>c</u> l <u>y</u> A <u>s</u> nP <u>h</u> eL <u>e</u> G <u>l</u> uT <u>r</u> V <u>a</u> l <u>l</u> eG <u>l</u>	338
D b	142	C <u>T</u> TAA <u>G</u> T <u>T</u> GCGAGGATGTGAAGATAATGGTCGAAGGATTCTCGAGTAGTTATTGAG	201

Qy	339	AanHisArgIleTyrGluGluProAlaThrThrPheGlnAlaPheGlnIleAlaLeuSer	358
Db	202	AANTTCGCCTCAACAGGAACCTCGAAGCACATTCATATGCTTCTTGATATCCCTGTGT	261

[illegible]

Qy	378	aAnSerGluLeuAlaPheLysCysSerArgValGluValAspTyrIleTyrSerIleLe	398
Db	322	GAACGAACACATTGGGTTTCAGTTGTAAAGGAAATGACGACACTCAGTGTCTCGAAGCT	381

Qy 398 userCyeMetLysSerLeuPheLeuGluHisThr-----ClnGlyLeuGlnPhe 414

Db 382 GGAAATTCGCAAGAGATGTTCTTACTTCACACGGACATCCAAAGAAACCTTCGTTTC 439

RESULT 11

BP133890
LOCUS

LOCUS BFL133890 2000 bp mRNA standard
DEFINITION BFL133890 MAT001 Nicotiana tabacum cDNA clone BY6204, mRNA sequence.
ACCESSION BFL133890
VERSION BFL133890.1 GI:32876775
KEYWORDS EST.

KEYWORDS	EST.	ORGANISM	SOURCE	Nicotiana tabacum (common tobacco)
				Nicotiana tabacum

ORGANISM	REFERENCE
<i>Nicotiana glauca</i>	(bases 1 to 568)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
asterids; lamids; Solanales; Solanaceae; Nicotiana.	

REFERENCE
I (pages 1 to 668)
AUTHORS
Matsuoka, K., Tashiro, G., Horiguchi, T., Demura, T. and Fukuda, H.
TITLE
Profiling growth-phase dependent gene expression of tobacco BY-2 cells by comprehensive microarray analysis


```

Db          729  TTGANANNATGAT---TCAACTGTGTCTCTTTGGAGGCNACCATGCTCAG 780
                                     |||||
RESULT 14   BH788688      560 bp   DNA      linear   GSS 28-MAR-2000
LOCUS      fzm021f001e02k0 fzm021f001e02 5', genomic survey sequence.
DEFINITION
ACCESSION  BH788688
VERSION    BH788688
KEYWORDS   GSS.
SOURCE     BH788688.1 GI:19797554
           Zea mays
           Zea mays
REFERENCE   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Zea.
AUTHORS    Budiman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.
TITLE      Genethresher: methylation filtered genomic sequences from maize
JOURNAL    Unpublished (2002)
COMMENT    Contact: Bedell JA
           Orion Genomics, LLC
           4041 Forest Park Ave, St. Louis, MO 63108, USA
           Tel: 314 615 6979
           Fax: 314 615 5975
           Email: jbedell@oriongenomics.com
           Plate: fzm021f001 row: e column: 02
           Seq primer: SK reverse
           Class: shotgun
           High quality sequence stop: 560.
FEATURES   Location/Qualifiers
           source          1..560
                        /organism="Zea mays"
                        /mol_type="genomic DNA"
                        /cultivar="Mol7"
                        /db_xref="taxon:4577"
                        /clone="fzm021f001e02"
                        /clone_lib="fzm021f001e02"
                        /note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA
                        end-repaired, size fractionated to enrich for the 0.5 to
                        5 kb fraction, ligated into HincII-digested pBCSK(-)
                        vector and electroporated into E. coli cells."
ORIGIN
Alignment Scores:
Pred. No.:      6.6e-11      Length:      560
Score:          198.00      Matches:     52
Percent Similarity: 61.98%   Conservative: 23
Best Local Similarity: 42.98% Mismatches:    38
Query Match:     6.16%     Indels:       8
DB:              8         Gaps:         2

SEQ3_K705_COPY_650_1270 (1-621) x BH788688 (1-560)

Qy    3  AspaAlaPheLLeuPheGlySerSerLeuAsnProSerHisAspValLysHisValGlu 22
Db    379  GATGCCATCATCATATATTGATGACTGGAATCCCATGATGACTTGAGAGCCCTTCAG 320
Qy    23  LysIleLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrSerValCys 42
Db    319  AGGGTCATGCATGGAGTACACAGCTGTGAACCTGTGCCTATTATTTTCGTTGTGACTCGTCTTTT 260
Qy    43  ThrValGluGluLysAlaLeuIleLeuAlaArgGlnAsnLysArgGlnAsn-LysAlaVa 62
Db    259  ACATGGAGGAAAGAGCTCTATTCTTCGAAGATGATGACCATATTTTGGATWAGTAATAT 200
Qy    62  lGluAsnLeuAsnArgSerLeuThrHisAlaLeuLeuMet-TripGlyAlaSerTyrLeuP 82
Db    199  TGTGAATATAACCCCTAGCTGAGCCATTGTGTGCCCCAGTTGGGCGCATCATTTCTCT 140
Qy    82  heAspLysLeuAspHisPhe-----HisSerSerGluThrProAspSerGlyValSerP 100
                                     |||||

```

Db 139 TCAATAGACTTGGAGGTTGCAGCAACATAGCTAC-----TCAAATGCTCTG 92
 Qy 100 heGluInSerIleMeTAspGlyValIleHisGluPheSerSerIleLeuSerSerIys 119
 Db 91 GTGATGAGCTTTTCATCGATAATGTAGATTGGAGTTTTCACAAATTTACTCAGCAG 33
 RESULT 15
 CA215647
 LOCUS SCLAD1140B12.9 AD1 Saccharum officinarum cDNA clone SCLAD1140B12
 DEFINITION 5', mRNA sequence.
 ACCESSION CA215647
 VERSION CA215647.1 GI:35264132
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
 complex.
 REFERENCE 1 (bases 1 to 669)
 AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
 TITLE The libraries that made SUCEST
 JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 COMMENT Contact: Arruda P
 Centro de Biologia Molecular e Engenhariaia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010. 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: parruda@unicamp.br
 Clone distribution: clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bccccenter.fcav.unesp.br
 Plate: 140 row: B column: 12
 Seq primer: T7 Promoter Primer.
 Location/Qualifiers
 source 1..669
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCLAD1140B12"
 /lab_host="DH10B"
 /clone_lib="AD1"
 /notes="Organ: seedlings inoculated with Gluconacetobacter
 diazotrophicans; Vector: pSport1; Site_1: SalI; Site_2:
 NotI; An unidirectional cDNA library generated from
 [seedlings inoculated with Gluconacetobacter
 diazotrophicans]. cDNA was prepared from polyA+ mRNA using
 SuperScript plasmid system kit (Invitrogen). The
 double-strand cDNAs were fractionated in a sepharose
 CL-2B 40cm-columns and fragments sizing between 0.8 and
 1.5 Kb were directionally cloned into the vector. Details
 of each source of RNA and library construction can be
 obtained at http://sucest.lad.ic.unicamp.br/public"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.86e-09 Length: 669
 Score: 186.00 Matches: 36
 Percent Similarity: 72.15% Conservative: 21
 Best Local Similarity: 45.57% Mismatches: 22
 Query Match: 5.78% Indels: 0
 DB: 6 Gaps: 0
 SEQ3_K705_COPY_650_1270 (1-621) x CA215647 (1-669)
 Qy 3 AspAlaPheIleLeuPheGlySerSerLeuAsnProSerHisAapValLysHisValGlu 22
 Db 431 GATTCCATTATCATTATGATGACTTGAATCCACTAATGATCGAAGGCTCTCGA 490
 Qy 23 LysAlaLysIleGluSerCysSerGluArgThrLysIlePheArgLeuTyrrSerValCys 42

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 19:44:24 ; Search time 61.6667 Seconds
(without alignments)
5156.776 Million cell updates/sec

Title: SEQ3_K705_COPY_650_1270
Perfect score: 3216
Sequence: 1 RADAFILFGSLNPSHDVHKH.....SGNNKHFGSSNISQNPADV 621

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3213	99.9	2001	2	Q9M658	Q9M658 arabidopsis
2	3213	99.9	2254	2	Q9LN02	Q9LN02 arabidopsis
3	3210	99.8	2001	2	Q9M659	Q9M659 arabidopsis
4	163.5	5.1	1461	2	O17909	O17909 caenorhabdi
5	163	5.1	1245	2	Q8IAN1	Q8IAN1 plasmodium
6	162.5	5.1	842	2	Q8ZU43	Q8ZU43 homo sapien
7	162.5	5.1	1739	1	CHD2_HUMAN	CHD2_HUMAN
8	161	5.0	4007	2	Q7SH24	Q7SH24 neurospora
9	160	5.0	1738	2	Q9S141	Q9S141 arabidopsis
10	157	4.9	1611	2	Q7RD43	Q7RD43 plasmodium
11	154.5	4.8	1268	2	Q7RG05	Q7RG05 plasmodium
12	154.5	4.8	2230	1	GOA4_HUMAN	GOA4_HUMAN
13	154	4.8	905	2	Q6BHF8	Q6BHF8 debaryomyce
14	154	4.8	2757	2	Q7RRR9	Q7RRR9 plasmodium
15	153	4.8	7210	2	Q9V7G8	Q9V7G8 drosophila
16	153	4.8	9270	2	Q8MLD9	Q8MLD9 drosophila
17	152.5	4.7	1330	2	Q22342	Q22342 caenorhabdi
18	152	4.7	1320	2	Q5JK25	Q5JK25 rattus norv
19	152	4.7	2067	2	Q9U308	Q9U308 plasmodium
20	152	4.7	2160	2	Q8IE50	Q8IE50 plasmodium
21	151.5	4.7	1033	2	Q676A5	Q676A5 oikopleura
22	151.5	4.7	1310	2	Q949K0	Q949K0 lycopersico
23	151.5	4.7	1931	2	Q9NCF9	Q9NCF9 drosophila
24	151.5	4.7	1931	2	Q9VKR9	Q9VKR9 drosophila
25	151	4.7	1500	2	Q75J40	Q75J40 oryza sativ
26	150	4.7	979	2	Q9EP71	Q9EP71 mus musculu
27	150	4.7	1939	1	MYH4_HUMAN	MYH4_HUMAN
28	150	4.7	2042	2	Q6BU09	Q6BU09 debaryomyce
29	150	4.7	8407	2	Q7RTW4	Q7RTW4 homo sapien
30	149.5	4.6	992	2	Q9NTH6	Q9NTH6 homo sapien
31	148	4.6	1456	2	Q9UPV0	Q9UPV0 homo sapien

32	148	4.6	1786	2	Q7ZT34	Q7ZT34 poephila gu
33	148	4.6	3187	2	Q63714	Q63714 rattus norv
34	147.5	4.6	1927	2	Q75WX9	Q75WX9 helicobacte
35	147.5	4.6	6761	2	Q6LEX8	Q6LEX8 plasmodium
36	147	4.6	962	2	Q96164	Q96164 plasmodium
37	146	4.5	628	2	Q6CAJ4	Q6CAJ4 yarrowia li
38	145.5	4.5	751	2	Q8URV5	Q8URV5 phthorimaea
39	145.5	4.5	1780	2	Q6FTH3	Q6FTH3 candida gia
40	145.5	4.5	1938	2	Q9PVE1	Q9PVE1 brachydanio
41	145	4.5	738	2	Q6CDT3	Q6CDT3 yarrowia li
42	145	4.5	1130	1	YL17_CAEEL	YL1102 caenorhabdi
43	145	4.5	1939	2	Q6DFQ6	Q6DFQ6 xenopus tro
44	145	4.5	8797	1	SNE1_HUMAN	Q8NF91 homo sapien
45	144	4.5	887	2	Q75AF5	Q75AF5 ashbya goss

ALIGNMENTS

RESULT 1

Q9M658 PRELIMINARY; PRT; 2001 AA.
AC Q9M658;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MOM.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20279299; PubMed=10821279; DOI=10.1038/35012108;
RA Amedeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RT "Disruption of the plant gene MOM releases transcriptional silencing
of methylated genes.";
RL Nature 405:203-206(2000).
DR EMBL; AF213628; AAF73381.1; -;
SQ SEQUENCE 2001 AA; 218567 MW; B9C85D4E62704441 CRC64;

Query Match	99.8%	Score	3213;	DB	2;	Length	2001;
Best Local Similarity	99.8%	Pred. No.	2.4e-170;				
Matches	620;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	RADAFILFGSLNPSHDVHKHVEKIKIESCSERTKIFRLYSVCTVEEKALILARONKRNK	60				
Db	650	RADAFILFGSLNPSHDVHKHVEKIKIESCSERTKIFRLYSVCTVEEKALILARONKRNK	709				
Qy	61	AVENLNARSLTHALIMWCASYLFDKLDHFHSSETPDGSGVSEQSIMDGVIFHSFSLSSKG	120				
Db	710	AVENLNARSLTHALIMWCASYLFDKLDHFHSSETPDGSGVSEQSIMDGVIFHSFSLSSKG	769				
Qy	121	GEENEVKLCILLAKHAQGYSSDSTLFGSDHIKLSDEESPNI FWSKLLGGKPNPWKYP	180				
Db	770	GEENEVKLCILLAKHAQGYSSDSTLFGSDHIKLSDEESPNI FWSKLLGGKPNPWKYP	829				
Qy	181	DTQRNKRQVQEGSEASPKTGCGNAKKRKKASDDVTPRVTDPPVDDDERKASGDH	240				
Db	830	DTQRNKRQVQEGSEASPKTGCGNAKKRKKASDDVTPRVTDPPVDDDERKASGDH	889				
Qy	241	MGALESFKVITLQSSCKSSGTDGTLGNDNDAFLYSGMSGSHISGIPEDMLASQDWKIPDES	300				
Db	890	MGALESFKVITLQSSCKSSGTDGTLGNDNDAFLYSGMSGSHISGIPEDMLASQDWKIPDES	949				
Qy	301	QRRLHTVLKPKMAKLCQVLHLSDACTSMWGNFLEYVYIENHRIYEEPAATTQAFQIALSWI	360				
Db	950	QRRLHTVLKPKMAKLCQVLHLSDACTSMWGNFLEYVYIENHRIYEEPAATTQAFQIALSWI	1009				
Qy	361	AALLVKQILSHKESLVANSELAFKCSRVEDVYIYISLSCMKSLFLEHTQGLQDFDCTGTN	420				
Db	1010	AALLVKQILSHKESLVANSELAFKCSRVEDVYIYISLSCMKSLFLEHTQGLQDFDCTGTN	1069				

QY 421 SKOSVSTKLVSLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 480
 DB 1070 SKOSVSTKLVSLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 1129
 QY 481 GIKKKYKQVQKLVQHEEKKMELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 540
 DB 1130 GIKKKYKQVQKLVQHEEKKMELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 1189
 QY 541 KLLDHNVERKFDIEKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLKVCVPI 600
 DB 1190 KLLDHNVERKFDIEKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLKVCVPI 1249
 QY 601 QSGNNKHFSGSSNISQNPV 621
 DB 1250 QSGNNKHFSGSSNISQNPV 1270

RESULT 2
 ID Q9LN02 PRELIMINARY; PRT; 2254 AA.
 AC Q9LN02;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE T6D22.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altafi H., Bei Q., Chin C., Chioi J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.,
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.,
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC026875; AAF79839.1; -.
 DR PIR; D86215; D86215.
 DR HSP; P00918; I8V3.
 DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
 DR GO; GO:0008270; P:zinc ion binding; IEA.
 DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
 DR InterPro; IPR001148; Euk_Coahnd
 DR InterPro; IPR011009; Kinase-like.
 DR Pfam; PF00194; Carb anhydrase; 1.
 DR ProDom; PD000865; Euk_Coahnd; 1.
 SQ SEQUENCE 2254 AA; 248157 MW; 64B80CF274791FF3 CRC64;

Query Match 99.9%; Score 3213; DB 2; Length 2254;
 Best Local Similarity 99.8%; Pred. No. 2.8e-170;
 Matches 620; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RADAFILFGSSLNPSHDVGHVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNK 60
 DB 883 RADAFILFGSSLNPSHDVGHVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNK 942
 QY 61 AVENLNRSULTHALLMWGASYLFDKLDHFHSSETPDGVSFEQSIMDGVHFEFSSILSKG 120

DB 943 AVENLNRSULTHALLMWGASYLFDKLDHFHSSETPDGVSFEQSIMDGVHFEFSSILSKG 1002
 QY 121 GEENEYKLCLEAKHAQCTYSSTDLFGEDHKLSDDEESPNIFWSKLLGGKPNMWPYS 180
 DB 1003 GEENEYKLCLEAKHAQCTYSSTDLFGEDHKLSDDEESPNIFWSKLLGGKPNMWPYS 1062
 QY 181 DTPQRNRKRVQVPEGEASPKTGDGGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH 240
 DB 1063 DTPQRNRKRVQVPEGEASPKTGDGGNAKRRKASDDVTDPRVTDPPVDDDERKASGKH 1122
 QY 241 MGALESFKVITLQSSCKSSGTDTLGDNAFGLYSMGSHISGIPEDMLASQDWGKIPDES 300
 DB 1123 MGALESFKVITLQSSCKSSGTDTLGDNAFGLYSMGSHISGIPEDMLASQDWGKIPDES 1182
 QY 301 QRRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRIYEEPAITFFQAFQIALSWI 360
 DB 1183 QRRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLEYVIENHRIYEEPAITFFQAFQIALSWI 1242
 QY 361 AALLVKQILSHKESLVANSELAFKCSRVEDVYIYISLCMKSLFLEHTQGLQDFCEGTN 420
 DB 1243 AALLVKQILSHKESLVANSELAFKCSRVEDVYIYISLCMKSLFLEHTQGLQDFCEGTN 1302
 QY 421 SKOSVSTKLVSLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 480
 DB 1303 SKOSVSTKLVSLSGATVRDEKINTKSMRNSSEDEECMTKRCSHYSTATRDIEKTIS 1362
 QY 481 GIKKKYKQVQKLVQHEEKKMELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 540
 DB 1363 GIKKKYKQVQKLVQHEEKKMELLNMYADKKQLETSKSVAAVIRITCSRTSTQVGD 1422
 QY 541 KLLDHNVERKFDIEKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLKVCVPI 600
 DB 1423 KLLDHNVERKFDIEKSEKNECLSKLEOMHEVAKKLADEACWINRIKSWAAKLKVCVPI 1482
 QY 601 QSGNNKHFSGSSNISQNPV 621
 DB 1483 QSGNNKHFSGSSNISQNPV 1503

RESULT 3
 ID Q9M659 PRELIMINARY; PRT; 2001 AA.
 AC Q9M659;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE MOM.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20279299; PubMed=10821279; DOI=10.1038/35012108;
 RA Aneudo P., Habu Y., Afsar K., Mittelsten Scheid C., Paszkowski J.;
 RT "Disruption of the plant gene MOM releases transcriptional silencing
 of methylated genes."
 RL Nature 405:203-206(2000).
 DR EMBL; AF213627; AAF73380.1; -.
 SQ SEQUENCE 2001 AA; 218585 MW; F6C5DAE897F2FE61 CRC64;

Query Match 99.8%; Score 3210; DB 2; Length 2001;
 Best Local Similarity 99.8%; Pred. No. 3.5e-170;
 Matches 620; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RADAFILFGSSLNPSHDVGHVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNK 60
 DB 650 RADAFILFGSSLNPSHDVGHVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNK 709
 QY 61 AVENLNRSULTHALLMWGASYLFDKLDHFHSSETPDGVSFEQSIMDGVHFEFSSILSKG 120
 DB 710 AVENLNRSULTHALLMWGASYLFDKLDHFHSSETPDGVSFEQSIMDGVHFEFSSILSKG 769

```

Qy 121 GEENEVKLCILLEAKHAQGYSSDSTLFGEDHIKLSDEESNIFWSKLLGGKPNMWKYP 180
Db 770 GEENEVKLCILLEAKHAQGYSSDSTLFGEDHIKLSDEESNIFWSKLLGGKPNMWKYP 829
Qy 181 DTPORNRKRVQYFGSSASPKTGGNAKKKASDDVTDPRVTDPPVDDDERKASGDH 240
Db 830 DTPORNRKRVQYFGSSASPKTGGNAKKKASDDVTDPRVTDPPVDDDERKASGDH 889
Qy 241 MGALSPKVIITLQSSCKSSGTDGLDGNDAFGLYSMGSHISGIPEDMLASODMGKIPDES 300
Db 890 MGALSPKVIITLQSSCKSSGTDGLDGNDAFGLYSMGSHISGIPEDMLASODMGKIPDES 949
Qy 301 QRRLLHTVLKPKMAKLCQVLHLSDACTSMVGNFLFYVIEHRIYIEPATTTFOAQIALSWI 360
Db 950 QRRLLHTVLKPKMAKLCQVLHLSDACTSMVGNFLFYVIEHRIYIEPATTTFOAQIALSWI 1009
Qy 361 AALAVKQILSHKESLVRANSELAPKCSRVEVDYIYISILSCMKSLFLEHTQGLQDCGTN 420
Db 1010 AALAVKQILSHKESLVRANSELAPKCSRVEVDYIYISILSCMKSLFLEHTQGLQDCGTN 1069
Qy 421 SKQSWSTKLWNESLSGATVDEKINTKSMRNSSEDEECMTKCSHYSTATRDIETIS 480
Db 1070 SKQSWSTKLWNESLSGATVDEKINTKSMRNSSEDEECMTKCSHYSTATRDIETIS 1129
Qy 481 GIKKKYKQVQKLVQEHKEMKELLNMYADKKQKLETSKSYEAAVIRITCSRTSTQVGD 540
Db 1130 GIKKKYKQVQKLVQEHKEMKELLNMYADKKQKLETSKSYEAAVIRITCSRTSTQVGD 1189
Qy 541 KLIDHNYERKFEDEIKSEKNECKLSLEQHEVAKKLADEACWINRIKSWAAKLKVCVPI 600
Db 1190 KLIDHNYERKFEDEIKSEKNECKLSLEQHEVAKKLADEACWINRIKSWAAKLKVCVPI 1249
Qy 601 QSGNNKHFGSSNISQNPV 621
Db 1250 QSGNNKHFGSSNISQNPV 1270

RESULT 4
O17909 PRELIMINARY; PRT; 1461 AA.
AC O17909;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein H06001.2.
GN ORFNames=H06001.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX BARLOW K.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Contains 2 chromo domains.
DR EMBL; Z92970; CAB07481.2; -.
DR PIR; T23056; T23056.
DR WormBase; WBGene00010369; H06001.2.
DR WormPep; H06001.2; CE32454.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.

```

```

DR GO; GO:0003682; F:chromatin binding; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006333; P:chromatin assembly or disassembly; IEA.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002464; DEAH_box_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; Chromo; 2.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEADC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00113; CHROMO_2; 2.
DR PROSITE; PS00690; DEAH ATP HELICASE; UNKNOWN_1.
KW Hypothetical protein, Nuclear protein.
SQ SEQUENCE 1461 AA; 168998 MW; 0DEA2BB5BAB91958 CRC64;

Query Match 5.1%; Score 163.5; DB 2; Length 1461;
Best Local Similarity 17.6%; Pred. No. 1.6;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

Qy 2 ADAPILFGSSINPSHDVHKIEKIESCSERTKIFRLYSVCTVEEKALILARQNKA 61
Db 799 ADTVIIIPDSWNPQDLQAMRAHRIGQTKTVNIYRLVTKGSVEE-----EI 845
Qy 62 VENLNRSLTHALLMWGASYLPDKLDH-----FHSSSETPSGVSFEQIMDV 108
Db 846 VERAKRLV-----LDHLVIORMDTTGTVLISKATAGSVFPDKQELSAI 891
Qy 109 IHFSSSI--LSSKGBENEV-----KLCLLEAKHAQGYSSDSTLFGEDHIKLSDEES 160
Db 892 L-KGAVELFKKEGEQEPEVDIDRILMGATREAEVEVMKENELLSFKYANFAIDEE 950
Qy 161 PNIP-----WSKLLGGKPNMWKYPSTDPQNRKRV-----Q 191
Db 951 KDIAAATDEWAAII-----PEEDNRNILEERMKELAEMLNAPRQRKQPIQP 997
Qy 192 YFGSEA-SPTGDDGNAKKKASDDVTDPRV-----TDPVDDDERK 234
Db 998 VVEDDDGDDDEEDDTGKKKKKAVGNFTIPEIKRPIKSRKPSMPLNRLEIAQDALE 1057
Qy 235 ASGKDHMGALSPKVIITLQSSCKSSGTDGLDGNDAFGLYSMGSHISGIPEDMLASODMG 294
Db 1058 EHSTDEMKKL---VESLSEACKKADE--FDSNEKNG-----DAG 1092
Qy 295 KIPDES-----RRLHTVLKPKMAKLCQVLHLSDACTSMVGNFLFYV 336
Db 1093 AAESEKDIERKPKFHTCDVNLKQIERSHAELKP---LHEILKSEETKTS----- 1139
Qy 337 IENHRIYEEPATTT-----FOAQIAL-----SWIAALLVKQI-LSHKESL 375
Db 1140 -----FKPPANAKLQKGVNDVMSRPDSDALLGVWKYGYGSGWEAIKMDPTLGLADKIFI 1193
Qy 376 VRANSELAFKCSRVEVDYIYISILSCMKSLFLEHTQGLQDCFGTNSKQSVVSTKLVNESL 435
Db 1194 KDKTKPKQGNLQVRVDYLLKLM-----SKDKVKTTE----- 1225
Qy 436 SGATVRDEKINTKSMRNSSEDEECMTKCSHYSTATRDIETISGKKYKQVQKLVQ 495
Db 1226 -----KKERKAKADDVPVGPPEKKRHTNNVPQEGE-----KKKEK----- 1262
Qy 496 EHEKMKELLNMYADKKQKLETSKSYEAAVIRITCSRTSTQVGDLLDHNRYERKFPDEIK 555
Db 1263 --BEKNSSSLK---DQALLSIDKSLYGLGALEDSASAKPFLECVKLCMPVHKYMKKLEAQ 1317
Qy 556 SEKNECKLS-----LEQHEVAKKLADEACWINRIKSWAAKLKVCVP 599
Db 1318 EAKQADEAKYFLRGDSFLENLETLKKPKPTNIRKWNVYLVFLCKFTLREP 1371

```

```
RESULT 5
Q8IAN1 PRELIMINARY; PRT; 1245 AA.
AC Q8IAN1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PF08_0127.
GN Names=PF08_0127;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
OX [1]
RN R1
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrall B.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844507; CAD51332.1; -
KW Hypothetical protein.
SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Query Match 5.1%; Score 163; DB 2; Length 1245;
Best Local Similarity 18.2%; Pred. No. 1.4;
Matches 126; Conservative 111; Mismatches 258; Indels 196; Gaps 27;

QY 12 LNP SHDVKHVEKIKIBSCSERTKIFRLY-----SVCTVEEKALILARQNRQKQAVENLN 66
Db 1 MNYDEDIKRKKKKTEKLNKTESLSLYESNNKILSIKKNLLKDKKNKMKIQKMKN-- 58
QY 67 RSLTHALLMGWASYLFFKLD-----HFHSSTPDSGVSFQSIWD----- 106
Db 59 -----KYDKDDGTIKHKKGNKIISDDKFDDEITNLNKKSKKCKMOTMVK 104
QY 107 -----GVHFEFSTLSSKSGEENEVKLCLLEAKHAQGTYSSTLTF 148
Db 105 KKKKKKQNYEKDKNVIGVMDEDEIID--NEERIENNENNVDHVPYTNDSK--- 159
QY 149 GEDHIKLSDESPNIFWFKLGGNPMWKYPSDTPQNRNRKVQYFEGSEASPTGCGGNA 208
Db 160 -----SDEYSSVNFMKLKN-----DNEREVNNKKKKKSTGDNHKNDDND 206
QY 209 KKKKASDDVTDPRVTPDPPVDDDERKASGDHMGALSPKVTILQSSCKSGTGTLDGN 268
Db 207 NDDNDDDDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDNDND 247
QY 269 DAFGLYSWGSIGIPEDMLASODWGI PDSEORRLHTVLKPKMAKLC-----OVLHLSDA 324
Db 248 DDNKKKKLVH-SGDQDDHVDVDDSSVQDSYQ---YALKDDMYKLSNENKEKIEPADL 303
QY 325 CTSWVGNF-----LEYVIENHR-----IYEEPATTFQAFIALSWIAALLVK 366
Db 304 MYSN-NNFFGGQLKDLDF-LENQNKKKGKRLSFIE-----QKENTTLGYVQNVKLL 356
QY 367 QILSHKESLVANSALAP-----KCSRVEDVYIYLSCKMSLPLEHTQGLQDFCFTNS 421
Db 357 NQNKAYNVQNSSKVFGRQKRGNGQGEDFVNKRL---VKRFLKREKIQPEMYDGE 413
QY 422 QSVV---STKLNESLSGATVRDEKINTKSMRNSSEDECMTEKCRSHYSTATRDTEKT 478
Db 414 ESGYIDNDNDNIINNNI-----LKKHKVDQTHNGYGEQNVYNNK-----EKK 457
QY 479 ISGIKKKYKQVQKLVQVHEEK-RWELNNMYADKKQKLETSKSVAAVIRITCSRTSTQV 537
Db 458 I--IKREHKEEDFTIYNFEIEIKWNL-----KSKMLHISED----- 493
QY 538 GDLKLLDHNVERKPEDEIKS-EKNECLSKLEQMHVAKKLAEDACWINRIKSWAAKLKV 596
Db 494 -----LLKNEKEKRDCLKIAQLKLLLSIENRKNYKKIKSKSYRYLRIKEKCEBEKI 549
QY 597 CVPTQSGN-----NKHFGSGSNISON 617
Db 550 MKKLYSEHPDLVKOLMNYEKEKAEKRNLLNN 580
```

```
RESULT 6
Q6ZU43 PRELIMINARY; PRT; 842 AA.
AC Q6ZU43;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ44007.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN R1
RP SEQUENCE FROM N.A.
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.,
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK125955; BAC8384.1; -
SQ SEQUENCE 842 AA; 94360 MW; F195761EFD08E33B CRC64;

Query Match 5.1%; Score 162.5; DB 2; Length 842;
Best Local Similarity 20.1%; Pred. No. 0.87;
Matches 123; Conservative 98; Mismatches 221; Indels 169; Gaps 26;

QY 122 ENEVKLCLLLBAKHAQGTYSSTLFGEDHILKLSDESPNIFWKLGG----- 171
Db 245 EKSEPKICRNLVTPKADPT-----GSEPAKASEKAPE---DVTDAEGESRREAA 293
QY 172 KNPWKY-----PSDTPQ-----RNRKRVQYFEGSEASPTGCGNKKKASDDVTD 220
Db 294 KEPKKASALEGSSDASQELSEHMKPQLSDSTASDPKSPHGLDFGRSISHELLD 353
QY 221 PRVTD-----PVDDERKASGDHMGALSPKVTILQSSCKSGTGTLDGN 253
Db 354 VDVLSPVLGACQAAQPLGIEDKDDSSQSDQLQSKGLSEYHRLSPPLPHEERAQ 413
QY 254 SSCGS-----SGTGTLD-----GNDAFLYSWGSISG-----IPEDMLA--- 289
Db 414 SPSRLATBEEPPQGPQEPQEWKEABELGEDSAASLSQLSLQRRSTEPVPAPPEQLSEAA 473
QY 290 -----SQDWGKIPDESORRLHTVLKPKM---AKLCO-----VLHL---SDACTSMVGNFL 333
Db 474 LKAMEAVQVLEQDQORHLLSEKQEKQQLREKLCQEBEEIILRLHQKQESLSLRELR 533
QY 334 EYVIENHRIYEEPATTFQAFIALSWIAALLVKQILSHKESLVANSALAFKCSRVEDVY 393
Db 534 QKAIE-----EEEARMEESQRLSWLRA-QVQSSSTQADEDDQIRAEQASLQKLRELE- 586
QY 394 IYLSILCMK-----SLPLEHTQGL-----QFDCFTNSKQSVST-----KLVSLSGA 438
Db 587 -----SQQAERASLSQKQRMLEQLKEIEA-SEKSEQAALNAAKEKALQQLREQLEGE 640
QY 439 TVDEKINTKSMNSSEDECMTEKCRSHYSTATRD-----EKTISGKKYKQVQKLV 494
Db 641 --RREAVATLEKHSAB-----LERLCSSLEAKHREVSVSLQKKIQEAQKEAQLQKCL 693
QY 495 QEHEEK-----KMLNNYADKKQKLE-----TSKSVEA 523
Db 694 GQVHRVHQKSYHVAGVEHLSLLREKQVEGEHERLDKMKEEHQVMAKAREQYEA 753
QY 524 AVRTICSTSTQVGDILKLDHNYERKFDIEKSEKNECLSKLSQMHVAKKLAEDACW 583
Db 754 EERKORAEILGLHTGELERLQRAHRELETVROEQHKRLERDRRRRREQERKLQDLELD 813
QY 584 INRIKSWAAKL 594
```



```

QY 272 GLY-----SMGSH-----ISGIPEDMLASQDWGKIPDESORRLHTVLKPKMAKLQCV 318
DB 1289 RFYRTVMKFGNHNQMACIABEVGVWE-----AAPBEAQVELPDA-----1328
QY 319 LHLSDACTSMV--GNF-----LEVVNIENHRIYEBPATTFO 351
DB 1329 --LIDGCKESVETGNFEPKGPVLDFGVPVKANELLKRVQGLQLLSKRIISRYNDPISQFR 1386
QY 352 AFQIALSWIA-----ALLVKQILSH-----KESLVRANSELAP--KCSRVVEV 391
DB 1387 -----VLSYLKPSNWSKCGWNQIDARLLGILYHGFNGWEKIRLDESGLTKKIAPVEL 1442
QY 392 DYIYSILSCWKSPLFLEHTQGLQDFP-----GTNSKOSVSTKLVLNLSLG---ATVRDE 443
DB 1443 QHHTETFLPAPNLKERATALLEMELAAAGGKNTAKASRNKSKVKVKNLQINQFKAPARDR 1502
QY 444 K-----INTK-----SNRSNSE-----DEE 458
DB 1503 RGKSGPANVLLSTKDGPRKTQKAPLVBEGEENSDDGEVYBPKEQKWMWECEDVLADE 1562
QY 459 CMTEKRCSHYSTATRDI--EKTISGIKKYK--KQVKQLVQBEHEKKMELLNNMYADKKQ 513
DB 1563 IKTLGRLQRLQTSADLPKEKVLKIRRYLEILGRIDAIVLEHEE-----DLKQDRM 1616
QY 514 KLETSKSVEAAVIRITCSRTSTQVGLKLLDHNRYERFDEIKSEKNECLSKLSQEMHEVAK 573
DB 1617 TMRL-----WNYVSTFSNLGDR-----LNQIYSKLGK 1643
QY 574 KLAEDAEACWINRIKSWAALKVCVPIQSGNNKHFSSNISQ 616
DB 1644 QKEEBEEGV-----GPHLNGSRNFOR 1665

RESULT 10
Q7RD43 ID Q7RD43 PRELIMINARY; PRT; 1611 AA.
AC Q7RD43;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY05582;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -i- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001787; EAA17620.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1611 AA; 188293 MW; CD0493CC4A9BC5D1 CRC64;

Query Match 4.98; Score 157; DB 2; Length 1611;
Best Local Similarity 21.08; Pred. No. 4.1;
Matches 134; Conservative 91; Mismatches 256; Indels 158; Gaps 26;

```

```

QY 25 KIBSCSERTKIFRLYSVCTVVEKALILAR--ONKRONKAVENLNRSLSLTHALLMGCASYLFD 83
DB 399 KYQINSKSKIIPKPDNSEKQFASLERLKNKWFANN--NENLKDSL-----D 448
QY 84 KLDFHSHSETPD--SGVSFEQSIQMDGVIHEFSSILSSKGGEEVNLCLLLEAKHAQGTY 141
DB 449 GIKKINKPTEEPKQINGSQSLGVNKSLELLKRLKKEGIYENGKSTLNFDPGK-----IT 504
QY 142 SSDSTLFGEDHIKLSDEESNIFWSLLGKGNPM--WKYPSPDTPQRRNRKVQYPEGSEAS 199
DB 505 SKDKT-----ESIK--KEEKENTFDSIRIKYENKINGSKMTNDVILNHTKNKSYDDLSSQ- 557
QY 200 PKTGDGNAKRKASDDVTPDPVPPVDDDERKASGKHDMGALESKPKVITLQSSCKS- 258
DB 558 -----NDDKSKINF-----IKDILPDDASMIHVRQORSIYSEKTEKTNQSNYNNY 604
QY 259 -SGTDTGLDGNAP--GLYSMGSHISGIPEDMLASQDWGKIPDESORRLHTVLKPKMAKL 315
DB 605 DCNIHGMDGDESTLTNNFRTVNKIYSGVKE--MLKQNMNKKNDISIYKDSNNAKVENMENY 663
QY 316 CQVLHLSDACTSMVGNFLEYVNIENHRIYEBPATTFOAFQIALSWIAALLVKQILSHKESL 375
DB 664 CESSIYSKNDNDKNGYIQ--LDNLKNFN-----TSNSVCLS-----NLKSENL 706
QY 376 VRANSELAFKCSRVVDYIYSILSCWKSPLFLEHTQGLQDFGFTNSKOSVSTKLVLNE 433
DB 707 ILAINEI--KQWMEKIDDKMSTIYSEKSELGIINNTK---DTGNGNNTGKIVVNGICN- 759
QY 434 SLSGATVRDEKIN-----TKSMRNSSEDEECHEKRC-----SHYSTATRD 474
DB 760 GLQGYSKDEKYNRMKFLIKNKKIKLKMSTYKGCYLNIRNLNLMKKNKIKFKFNKL 819
QY 475 IEKTISGIKKYKQVKQLVQBEHEKKMELLNNMYADKKQKLETSKSVEAAVIRITCSRTS 534
DB 820 LKQYDNLKSKYKNVLKLC-----KKIFLLKYTTYK-----MHEVAKKLADEA 581
QY 535 TVQVDLKLKLDHNRYERFDEIKSEKNECLSKLEQ-----NRIKSWAALKV 596
DB 852 ----DRSNYDKKYENWMRHQLSEKNLNLIEEKYEQENININMNLMSDELKQLQENE 907
QY 582 CWI-----NRIKSWAALKV 596
DB 908 SIISNNNSYKTEVETINSKYQLNDNFNIKSEHEKLKI 946

RESULT 11
Q7RG05 ID Q7RG05 PRELIMINARY; PRT; 1268 AA.
AC Q7RG05;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 10.
GN Name=PY04546;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria

```

```
RT parasite Plasmodium yoelii yoelii." ;
RL Nature 419:512-519(2002).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAL01001392; EAA16423.1; -.
DR GO: GO:0003743; F:translation initiation factor activity; IEA.
DR InterPro: IPR000717; PCI.
DR Initiation factor.
KW SEQUENCE 1268 AA; 153041 MW; DE3D62D55C47B91C CRC64;

Query Match      4.8%; Score 154.5; DB 2; Length 1268;
Best Local Similarity 20.1%; Pred. No. 4.1;
Matches 150; Conservative 111; Mismatches 253; Indels 233; Gaps 36;

QY 5 FILGSSLPNSHDVHKVEKIKIESCSERTKIFRLYSVCT-----VEKALILARONKRON 59
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 FCYLEKSLVKGDLGHQYRI-----ICQGNIASLGKVTDFRDKAEK-VRLAKENNNIN 112
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 KAVENLNRSILTHALLMWGASYLFDKLDHFHSSSETPDGVSF-EQSIMDGVIIHFSSILSS 118
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 113 K-----DKI-----EQEESNVDFTEKILMNSLDMITGKYER 144
QY 119 KGGEENEVKLC-----LLLEAKVA-----QGTY-----SSDSTLFGEDHKLSD-----EESPN 162
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 145 K--LQNAKIKTCMETYKIMLEILRATPKLEKAYHDTAKKALLFCENKRLTEFKKLSDLLR 202
QY 163 IFWSKLLGGKNP-----MWKYPSPDTPQ--RNRKRVOYF 193
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 203 NHYNLILRGHKPEYQSLKIEVHLETKIMQLETACELGHWKASNAIEDIYNLMHHEYF 262
QY 194 -----EGSEASPKTGGGNAGKRRKASDDVTPRVTDPVDDDERKASGRDHMGALSPK 248
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 YKTLKSGAMESDMNNADQNSKKSSD-----KDEKDTNEATTSGGVASTS 310
QY 249 VITLQSSCK-----SSGTDGTLGNDAFGLYSNGSHIS----- 281
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 311 AIASQNKQENLKKWIAFTFEKLAIDLIFVYSESLFHLGLAWLKTCFHLNYGIEISERKK 370
QY 282 -----GIPE-DMLASQDWGKIPDSORLHLVLPKMAKLCOVHLHSLDACTSMV 329
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 371 FICTKAVLAVLSIPRIGKKNEDYAKI--FEAKKMSQLLGHSTVPVKESLNG-----LKV 425
QY 330 GNFEVYVIEN-HRIYEPATTFQAFQIALSWIAL-----LVKQILSHKESL 375
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 RNILNADENAKQLSYLVENKFTPLSLCEVLLEKLETEHNIYINKIKVIFHKLIJ 485
QY 376 VRANSELAFCRSRVVDY-IYSILS-----CMKSLFLEHTQGLQFDC-----F 417
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 486 -----QLSKVYNYISIDYFENICSEHFISWDAETMLVDLVYQKELEMRIDLTKRAIYF 540
QY 418 GTNSKQSVSTKLAVNESL-----SGATVREKINTKSMRNSSEDEECWTE 462
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 541 G--DKNVNRSLKIDSLINWYLNGLRMINESISIEHVENVTNKLNYEHFNLID 598
QY 463 KRCSHYST-ATRDIEKTISG-----IKKKYKQVQKLVQSHBEKQWELLMYADKK 512
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 599 KGNKHSINYATADIEDIVENEIIDDKPIDDEKILKKIYDKIDESH--KLQLLSEHNKK 656
QY 513 QK--LETSKVEAAVIRITCSRTSTQVGDLLKLDHNYERFDETKSEKNECLSKLEQWHE 570
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 657 RKEMLRKQEQEQQAQLKIRMEK-----RLLEKLEKREKLEAKKGEKLRIRKE--K 706
QY 571 VAKKCLAEDACWNRITKSWAAKLKVC 597
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 707 IKKKEAAEQ--MLKEIK-----KJC 725

RESULT 12
GOA4 HUMAN
ID GOA4 HUMAN STANDARD; PRT: 2230 AA.
AC Q13439; Q13270; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
```

DR DR Genew: HGNC:4427; GOLGA4.
DR MIN; 502509; --
DR MIM; 270150; --
DR GO; GO:0005802; C:Golgi trans face; TAS.
DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR InterPro; IPR000237; GRIP.
DR Pfam; PF01465; GRIP; 1.
DR PROSITE; PS00913; GRIP; 1.
KW 3D-structure; Alternative splicing; Antigen; Coiled coil; Golgi stack.
FT DOMAIN 133 2185 Coiled coil (Potential).
FT DOMAIN 252 2096 Glu-rich.
FT DOMAIN 2168 2215 GRIP.
FT VARSPLIC 2154 2185 TPYKGNLYHTDVSILGCEPTEFEYLRKVLPEY -> HLTKV
FT AICITRMHSHLENLPNLSICEKFLSI (in isoform
FT 2).
FT /FTid=VSP_004272.
FT Missing (in isoform 2).
FT /FTid=VSP_004273.
FT Missing (in isoform 3).
FT /FTid=VSP_004274.
FT /FTid=VSP_004275.
FT FTSPRSGLF -> SWLRSSS (in isoform 4).
FT /FTid=VSP_004275.
FT R -> K (in Ref. 3).
FT CONFLICT 188 188 Y -> H (in Ref. 3).
FT CONFLICT 220 220 Y -> H (in Ref. 3).
FT CONFLICT 276 276 T -> A (in Ref. 3).
FT CONFLICT 584 584 K -> E (in Ref. 3).
FT CONFLICT 628 628 T -> A (in Ref. 3).
FT CONFLICT 630 630 K -> E (in Ref. 3).
FT CONFLICT 682 682 K -> N (in Ref. 3).
FT SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;
Query Match 4.8%; Score 154.5; DB 1; Length 2230;
Best Local Similarity 19.7%; Pred. No. 8.5;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;
QY 13 NPSHDVKHVEKI-KIESCSB-----RTKIFRLYSVCTVEEKALILARQNKQKAVEN 64
DB 1552 NQKDIIEHKLVLQKQFQELGEKDNRV-----EABEKLITLENQVYSMKAELET 1603
QY 65 LNRSLTHALLMWGA-----SYLFKDLHFHSSPTDGVSPFQSGIMDGVHFEFSSILSSK 119
DB 1604 KKLEHVNLSVSKSEBELKALEDRLESNAKLAELKRAEKOKIA-AIKQLLSQMEK 1662
QY 120 GBEENEVKLLLEAKHAQGYTSSDSTLFGEDHKLSD-EESPNIFWSKLLGGQNPWKY 178
DB 1663 -----BEQYKKGTEHSEL-----NTKLOEREREVHILEEKLKSVESQ--- 1702
QY 179 PSDT---PQRNRKRVQYFEGSEASPKTGDGNAKRRKKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPSAKNVAAYTEQEEADSQ-----GCQVKTYEEKISVLQRLNTEKEKLQVRGQ 1757
QY 230 DDERKASGKDHMGALSPKVTITLQSSCKSGTGTDLG-----NDAFGLYSMGSHI 280
DB 1758 EKEETVSSHPEMRQYQERLIKLEHAEAKHQEDSMIGLQBELEEKKKYSLI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKIPESQRRLHTVLKPKNAKLCQVHLIS-----DACTSMVGNFLE 334
DB 1817 EKEGKNNIOAKQNLNVFDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
QY 335 YVIENHRI-YEEPATTQAQIALSWI-----AALLVKQILSHKES 374
DB 1865 RQKEVHRVHEEULTSKYEKLQ-ALQQMDGRNKPTELEENTEKSKSHLVQPKLLSMEA 1923
QY 375 LVANSELAFKCSRVED-----YIYSILSCMKSLFLEHTQGLQFDCFGTNSQSVVST 428
DB 1924 ---QHNDFLEKLAGAEKQKGLKEIVRLQKDLRLMRKEHQEQL-----ILKK 1969
QY 429 KLVNESLGATVRDEKINTK-----SMRNSDEECWTEKRCSHYSTATRDIEKTISGIKK 484
DB 1970 EYDQE-----REEKIQEQEDELKXNSTLQKLMREFN-TOLAQKEQELEMTIKETIN 2021
QY 485 KYKQOVOKLQVEHEEKKMELLNVADKKQKLETSKSVAAVIRTCSTSTQVGDKL-- 542

Db 2022 KAQVEAEELLESHEQETNQLKKIAEKDDDLKRTAKRYEEILDAREEBMTAKVRDLQTL 2081
Qy 543 --LDHYERKPFDEIKSEKNECLKSLEQMHVEYAK-----KLAEDACWNRIRKSWAAK 593
Db 2082 EELQKYYQKLEQENPNNDNVITMELQTLQAKTTLISDSKLKEQB--PREQIHNLDR 2139
Qy 594 LK-----VCVPQSQGNKH 607
Db 2140 LKYEKNVYATVGTGYKGNLYH 2163
RESULT 13
Q6BHP8 PRELIMINARY; PRT; 905 AA.
ID Q6BHP8
AC Q6BHP8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA2768|IPF17011 Candida albicans IPF17011.
GN ORFNames=DEHA0G201199;
OS Debaryomyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marek C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigie M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolaki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RL Nature 430:35-44 (2004).
RT "Genome evolution in yeasts."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382139; CAG90870.1; --
DR InterPro; IPR002017; Spectrin.
SQ SEQUENCE 905 AA; 103103 MW; F7B4B88E0D525F0C4 CRC64;
Query Match 4.8%; Score 154; DB 2; Length 905;
Best Local Similarity 19.0%; Pred. No. 2.9;
Matches 131; Conservative 106; Mismatches 240; Indels 214; Gaps 29;
QY 52 ARQNKQKQKAVENLNRSLTHALLMWGASYLFD-----KLDHFFHSSPTD 95
DB 48 AKQDARNTVHETPGLTATPLTKAEPLYSVKNRIPLTSPRATNPDLNLSKSKTK- 106
QY 96 SGVSFEQSIQMDGVIHEFSSILSKSGGEENEVKLLLEAKHAQGYTSSDSTLFGEDHKL 155
DB 107 -----LDARLYELKS--SSRHGSPSN-----KENMHSNGKSTQASQSSDVNSSI 149
QY 156 SDESPNIFWSKLLGGKNPWKYPS--DTPQRNRKRVQYFEGSEASPKTGDGNAKRRKK 213
DB 150 SH-----KSPFESSPKIDTPIK-QINSSFAKNVKSPPKSPSKADKRS- 193
QY 214 ASDVDTPRVTPDPPVDDDERKASGKDHMGALSPKVTITLQSSCKSS-----GTGD 263
DB 194 -----IEDDSVIENSRKIA-----KVITNANESEVSHKSDDEHLIOID 233

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA LaBko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacieb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Friese E., Wheeler D.A., Lewis S.B., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berwan B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; AAF58087.2; -;
DR HGSP; P56276; 1PHG.
DR FlyBase; FBgn0013988; Strn-Mlck.
DR GO; GO:0004685; F:calcium- and calmodulin-dependent protein k. . .; IDA.
DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; ig; 3.
DR SMART; SM00408; IGc2; 8.
DR PROSITE; PS00835; IG LIKE; 19.
DR PROSITE; PS00290; IG_MHC; UNKNOWN 1.
DR SEQUENCE 7210 AA; 808985 MW; 5B2E7395C07140D2 CRC64;

Query Match 4.8%; Score 153; DB 2; Length 7210;
Best Local Similarity 20.1%; Pred. No. 46;
Matches 138; Conservative 92; Mismatches 239; Indels 218; Gaps 30;
Qy 3 DAFILGSSLNPSHDVHVEKIKIESCSERTKIFRLYS----- 40
Db 3377 DIFVLDLFDL---DETEVLKPKIENI---KTTLSLDYDIYEKKDGPLLTAVINKINVVS 3431
Qy 41 --VCTVEEKALILARQNKRONKAVENLNRLSLTHALLMWGASYLPDKLDHPSSTPDSGV 98
Db 3432 QHILTIIEVKQLTENHDKQKVSNAEAD-----NPADEKEESQKEIKUSEA 3481
Qy 99 SFEOISIMDVITHEPSSILSKGGEENVKGLLEAKHAQCTYSSTDLFGEDHFKLSIE 158
Db 3482 K-----HKSKVSEKKSIEEEK-----LEDKKEKQTESA-----IDE 3513
Qy 159 ESPNIFWSKLLGGKQPMWKYPSTDPQRNKRQVQFEGSEASPK----- 201
Db 3514 KSQKAEVSEIVSE-----KITDEKAQESQKK--EVKGEAKPKKAKVLEKKSIEEEKLE 3566
Qy 202 -----TGDCGNNAKRRKKASDDVTPDPPVDDDERKASQKHMGALSPKVTTLQS-- 254
Db 3567 KKEKQTESAIDEKSQKAEVSEIVSEKITDEKAQESQKK--EVKDEAKPKKAKVLEKKSIE 3625
Qy 255 -----SCKSGTGTDTLGNDAFGLYSMGSHISGIPEDMLASQDWGKIPDESQRR--LHTV 307
Db 3626 EAKLEDKKEKQTQDTSIDEK-----SQKAEVSEIVSEKITDEK-----AQESQKEEVKDE 3675
Qy 308 LKPKMAKLCQVHLSDACTSMVGNFLFYIENRIYEEPATTPQAFQIALSWIAALLVKQ 367
Db 3676 AKPKKAK-----VLEKKSIEE-----K 3693
Qy 368 ILSHKBSLVRANSELAPKCSRVEVDYIYISILSCMKSLFLEHTQGLQ----FDCFGTNSKQ 423
Db 3694 LEDKKEK--QTESAIDEKSQKAEVSEIVS-----EKITDEKAQESQKEVKSQKAEVSEI 3746
Qy 424 SVVSTKLVBESL---SGATVRDEKINTKSMRNSSEDEECMKRCSHYSTATRDIETIS 480
Db 3747 KVLEKKSIEEEKLEDKKEKQTQDTSIDEKSQK--AEVSEIVSEKITDEKAQESQKEVKSQK 3804
Qy 481 GIKKYKQVQKLVQEHHEKMKMLNNYADKKOKLETSS-----KSVAAVIRITCSRTSTQ 536
Db 3805 EAKPKKAKVLEK--KSTIEEKELE-----DKGKQTESAIDEKSQKAEVSEI-----VSEK 3852
Qy 537 VGLDKLID--HNYERKFTDEIKSEKNECI--KSLE-----OMHE 570
Db 3853 ITDEKAQESQKEVKSQKAEVSEIVSEKITDEKAQESQKEVKSQKAEVSEI----- 3912
Qy 571 VAKKKLADEA--CWINRIKSWAAK 595
Db 3913 IVSEKITDEKAQESQMEVKSQKAEVSEIVSEKITDEKAQESQMEVKSQKAEVSEI 3939

Search completed: July 12, 2005, 01:01:40
Job time : 64.6667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005, Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2529.46 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929C-1_COPY_1_500
Perfect score: 500
Sequence: 1 aatatttaagttgggttat.....tttggtctatttaattctca 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10329	6 AX063582	AX063582 Sequence
2	500	100.0	10329	8 AF213628	Arabidops
3	500	100.0	120965	8 AC026875	AC026875 Genomic s
4	58	11.6	696	9 HSA333293	AJ333293 Homo sapi
5	57.6	11.5	170606	2 BX957335	BX957335 Danio rer
6	57.2	11.4	150671	2 CR847829	CR847829 Danio rer
7	57.2	11.4	184431	2 CR628331	CR628331 Danio rer
8	56.4	11.3	527	3 AY672413	AY672413 Meloidogy
9	56.2	11.2	162190	2 CR627484	CR627484 Danio rer
10	55.6	11.1	144022	5 BX088594	BX088594 Zebrafish
11	55.4	11.1	228431	2 CR788237	CR788237 Danio rer
12	55.2	11.0	110000	2 PFMAL13_17	Continuation (18 o
13	54.6	10.9	50000	6 AX392735	AX392735 Sequence
14	54.6	10.9	150199	9 AC087568	AC087568 Pan trogl
15	54.6	10.9	232392	14 AF250284	AF250284 Amsacta m
16	54	10.8	110000	2 PFMAL13_03	Continuation (4 of
17	54	10.8	139068	2 CR407548	CR407548 Danio rer
18	54	10.8	177623	2 CR394571	CR394571 Danio rer
19	53.8	10.8	110000	2 PFMAL13_21	Continuation (22 o

20	53.6	10.7	254050	3 PFA929358	AL929358 Plasmodiu
c 21	53.4	10.7	200796	9 AC151667	AC151667 Pan trogl
22	53.2	10.6	135121	9 AC069525	AC069525 Homo sapi
23	53.2	10.6	145650	2 CR762412	CR762412 Danio rer
24	53.2	10.6	337203	3 CR382401	CR382401 Plasmodiu
25	53	10.6	138769	2 AL139261	AL139261 Homo sapi
c 26	53	10.6	135717	5 BX321868	BX321868 Zebrafish
c 27	53	10.6	251762	3 AE014851	AE014851 Plasmodiu
c 28	53	10.6	256172	2 AC005139	AC005139 Plasmodiu
c 29	53	10.6	310779	2 AC005140	AC005140 Plasmodiu
30	53	10.6	333321	3 AC116986	AC116986 Dictyoste
31	52.8	10.6	125632	2 CR753868	CR753868 Danio rer
c 32	52.8	10.6	163167	9 AC099784	AC099784 Homo sapi
33	52.8	10.6	187415	2 CR759968	CR759968 Danio rer
34	52.8	10.6	217621	2 BX901924	BX901924 Danio rer
35	52.6	10.5	34548	6 AX349035	AX349035 Sequence
36	52.6	10.5	179759	9 AC084355	AC084355 Homo sapi
c 37	52.4	10.5	337203	3 CR382401	CR382401 Plasmodiu
c 38	52.2	10.4	11929	3 AE001431	AE001431 Plasmodiu
c 39	52.2	10.4	110000	3 PFMAL1P2_0	AL031745 Plasmodiu
c 40	52.2	10.4	155711	3 VYIVD10	AL360354 Plasmodiu
c 41	52.2	10.4	163930	2 CR847885	CR847885 Danio rer
c 42	52.2	10.4	169794	2 AC004688	AC004688 Plasmodiu
c 43	52.2	10.4	196149	2 AC004709	AC004709 Plasmodiu
c 44	52.2	10.4	252650	3 AE014847	AE014847 Plasmodiu
45	52	10.4	96799	9 AC104647	AC104647 Homo sapi

ALIGNMENTS

RESULT 1
AX063582
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

AX063582 10329 bp DNA linear PAT 24-JAN-2001
Sequence 1 from Patent WO0100801.
AX063582
AX063582.1 GI:12541306
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
1
Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
Gene involved in epigenetic gene silencing
Patent: WO 0100801-A.1 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

Location/Qualifiers
1. .10329
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

Query Match 100.0%; Score 500; DB 6; Length 10329;
Best Local Similarity 100.0%; Pred. No. 2e-77;

Handwritten signature

Db 241 GCAGTGGTTACAGTCTACTGCGAGGCCAAACTCGTAATTAACGTTTGAAGTTTAGAGA 300
Qy 301 AATATTTTGTATGATGAGTACACAAATCAAGATGATAGTGTGTTAATCACTGTAAATATGT 360
Db 301 AATATTTTGTATGATGAGTACACAAATCAAGATGATAGTGTGTTAATCACTGTAAATATGT 360
Qy 361 TGATGTAATGATGATGAGTACACAAATCAAGATGATAGTGTGTTAATCACTGTAAATATGT 420
Db 361 TGATGTAATGATGAGTACACAAATCAAGATGATAGTGTGTTAATCACTGTAAATATGT 420
Qy 421 TTTGTTTATCATTTTGAATACACGAAGAGCTCAGTTATATTTTATATTTGATATATGAA 480
Db 421 TTTGTTTATCATTTTGAATACACGAAGAGCTCAGTTATATTTTATATTTGATATATGAA 480
Qy 481 TTTGTTTATCATTTTGAATACACGAAGAGCTCAGTTATATTTTATATTTGATATATGAA 500
Db 481 TTTGTTTATCATTTTGAATACACGAAGAGCTCAGTTATATTTTATATTTGATATATGAA 500

AC026875 120965 bp DNA linear PLN 28-JUN-2000
Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
1, complete sequence.
AC026875
AC026875.4 GI:7651711
HTG.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Brassicales; Brassicaceae; Arabidopses;
1 (bases 1 to 120965)
Shinn, P., Brooks, S., Buehler, E., Chao, O., Johnson-Hopson, C.,
Khan, S., Kim, C., Altafi, H., Bei, Q., Chin, C., Chiou, J., Choi, E.,
Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, F.,
Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,
Mukharshy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G.,
Federapfel, N.A., Theologis, A. and Ecker, J.R.
Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
1
Unpublished
2 (bases 1 to 120965)
Ecker, J.R.
Direct Submission
Submitted (25-MAR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 120965)
Ecker, J.R.
Direct Submission
Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 120965)
Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, O.,
Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C.,
Chiou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N.,
Howing, B., Koo, F., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,
Liu, S., Mukharshy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,
Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M.,
Yu, G., Davis, R., Federapfel, N., Theologis, A. and Ecker, J.R.
Direct Submission
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Apr 27, 2000 this sequence version replaced gi:7543638.
Location/Qualifiers
1. 120965
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

chromosome="1"
/clone="T6D22"
join(410..559,596..758,828..865,1681..1731,1903..2253,
2562..2999,3096..3146,3240..3322,3493..3659,3781..3819,
3992..4021,4069..4246,4597..4703,4824..4952,5058..5122,
5239..5394,5549..5709,5812..5843,6103..6266,6382..6461,
6658..6723,6881..7170,7235..7264,7463..7578,7663..7760,
7939..7999,8364..8440,8552..8623,8702..8778,8863..8978,
9337..9439,9593..9751,9883..10281)
/notes="similar to AT.1.24-9 gene product gb|AAB07881.1;
similar to EST emb|Z35021.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.1"
/protein_id="AAF79821.1"
/db_xref="GI:8778822"
/translation="MIGHNFCGRNRILDSLPKTKNSLSOTIVAVPKISFVNRCS
TQVTKICVTKPSSSPNOADKCGKRYVAASRVCGVLLRLNKLQKPRD
LITNKVILTOGIDGFMVFEFDFESFDLASKFIDFCLLRWSPPHNTSINSVVPV
SQLVSDARRRPNLVAALISLYFYVVRKLSHVLMSLPITHTHFAFSNIHLATSRTS
SLRSSPFTLARSILCFNSSIRFLSSDMPKQKRDHAEQKQKPKYMGVSSSSAT
SATVAEAVNNQFGLSLKESNTAPVLPQSTTNRVONLVNPKPKYGTVSSSSAT
EVGKTSAVSQIGSGGDTKVLNLSKIFGGLNLEKFSVDKSTYCHQITRATFPKFE
KTQETRTMIENVSGKGLATLEVLKHSGLFMYAGHKGAYAKSFGNFIIVSYAQG
VLLADLDVTMDKIPHHITVVRQIYFKHGFDEANAFIYSFSLMLRGLGLQK
KQHGQRPLDDYDNSKAVQLKGNVDLHLAVVAVTAVTELGNGKPFYSTSEISFCRK
WRLPTNHWLFTSRKTSVTFPAFDALCEGIATSVCRALDEVADIICAAHRSNEKQ
GEITGLVARISSQSDRMENVRDHPHPPCDGANLDGLSLREICAAHRSNEKQ
RALLRSVGSFCPSDEWDESHPSADKSVITKLSQPADYSTKLSQEMVLEKME
KRUPAFCVHNFRABDISPNLFYKLVVHVSDSGFRYHKEMRHPSLWPLRYGF
FDINILFKNGAAYKTYIRQMKITNGTSGDKELCKMLDEWAIYIRKCNQDLS
IRNLSILFKNGAAYKTYIRQMKITNGTSGDKELCKMLDEWAIYIRKCNQDLS
SSYLSAEPFLEQYAKRSPKHLISAGNLVTFDLAIVDGLDEEDLVKQGV
TPATPAPKAEVQKGLIYFPVSVIELMDGSGIPGSAKSALCKELLNAPGFGD
RPVHTLMDGLVCKYKPVKADERRKQSLMLADKAPNEDVWQIDCMCRRTASAV
RNFSELIERFGLSILKMLKSDRTPDPVNVKGLDKESNAGVLLMPLHYEGKN
THAAGTWEKQLRDLTVANSEYLSIQVFESMVHQVREELTKIAGDKYKPSSEK
KYSIVFAINPATOVHSLLEKLAANPTMRPLEGKKSIQEKLSHVTLLAKHS
HGATVAVSQHLNRVPELTETIYNDKMAALTAVHGSVDGETVWSKNWPHVTLWT
HGVATAKEANTLPQVLEKASRLVIDPPVVISGPLEFF"
join(11454..11515,12009..12885,12925..13073,14200..14235,
14754..14777,14903..15370,15466..16353)
/notes="putative elongation factor 1-a
emb|Z17990.1, gb|AW220334.1, gb|AW622026.1,
dbj|AV439770.1, gb|T45275.1, and gb|AW267999.1,
gb|T04414.1, gb|AW221780.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.2"
/protein_id="AAF79822.1"
/db_xref="GI:8778823"
/translation="MGKRPKPHINIVIGHVDSKSTTTHLIYKLGIDIKRVIERPEK
EAAEMKRSFKAWLDKLAEREGITIDIALKPFETTKYCTICACDGRHDFIKNM
ITGTSQADCAVLIIDTGTGFEAGISQDQREHALLATFLGVKQMICCNKQDITTP
KYSKARVDEIKVSVYKGVNPKIPFVPSGEGDNMIERNLWPKYVTPAGLTLE
ADQINPEKPSDKPLRLQDVYKIGGIVPVGVEGTGMIKPGMVVTPAGLTLE
VKSVMHSHLEALPDNVGNFVNKAVDKRGVANSKDDPAKGAANFTSQVIL
MNHGQINGYAEVLDCHTSHIAVSEILTKIDRRSGKEIEKPEKPKNGDAGMVKM
TPTKPMVTFSEYFPLGRFAVDMRQTVAVGKSDKQDTPKATKAAVKGESK
DSYFTIYVYLVACLSCFSPSBCSRNWLDRRWIYTHILPHHSPLECDIVC
MIHNFVLTMGKEPHINIVIGHVDSKSTTTHLIYKLGIDIKRVIERPEKAAE
MKRSFKYAVLDKLAEREGITIDIALKPFETTKYCTICACDGRHDFIKNM
SQDCAVLIIDTGTGFEAGISQDQREHALLATFLGVKQMICCNKQDITTP
ARDEIIVKSVSYKGVNPKIPFVPSGEGDNMIERNLWPKYVTPAGLTLEADQ
INPEKPSDKPLRLQDVYKIGGIVPVGVEGTGMIKPGMVVTPAGLTLEADQ
EMHSHLEALPDNVGNFVNKAVDKRGVANSKDDPAKGAANFTSQVILMNH
GOINGYAEVLDCHTSHIAVSEILTKIDRRSGKEIEKPEKPKNGDAGMVKMTPK
PMVTFSEYFPLGRFAVDMRQTVAVGKSDKQDTPKATKAAVKGAK"
complete(join(19245..20132,20226..20687))
/notes="putative elongation factor 1-a
gb|AAF13128.1|AF124801.1; similar to ESTs gb|AW034714.1,
gb|AAF13128.1|AF124801.1; similar to ESTs gb|AW034714.1,

gb|AI012185.1, emb|Z17990.1, gb|AW267999.1, gb|N96009.1,
dbj|AV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"
/translation="MGKEKPHINIVIGHVDSKSTTTGHLIYKLGIDKRVIERFEK
EAAEMKRGPKYAWLDKKAERGIITIDIAWKPEFTTKYCTVIDAPGHRDFIKNM
ITGTSQADCAVLIIDSTTGGFEAGIKSGDGTREHALLAFTLGVKOMICCNKMDATTP
KYSAKYDEIIEKVSYLKKVGNPDKIPVPIVSGEGDNMIERSNLDWYKGPITLLE
ALDOINEPRSPKRLPLQDYKIGIGITVPGRVETGMIKPGMVWTFAPRGLTTE
VKSVEHSHSLLEALPDNDVGNVAVKLDKRGVVASNSKDDPKAGKAAFNFTSVQII
MNRPGIIGNGYAEVLDSHIAVFSEILTKIDRRSGKEIEKEKPELKNKRGAGWYK
TPTPKMVETFSYPLGRFAVRMDRQTVAVGVIKSVKDDKDPGAKVTKAAVKKGAK"
/complement(20787..20857)
/product="tRNA-Lys"
/note="codon recognized: AAA; putative tRNA-Lys"
/evidence=not experimental
complement(join(21956..22189,22535..22600,22713..22929,
23041..23126))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translations="MCLNRLPSATPIPRABERECCSMFGFRNHOLLSSPLCKRPA
IIMKGGSGGSGSGPTAAANAALOKKALLQKVALQVETDITSVNDNFTQIVNVSrv
SDPKNQSTRTYMEWRKARWQAASLLKLSLSELQKALFSGFASLNHVEORIEEF
DQBAEKNTLLARIADDSANLKELESYYSSAQLTLDI"
join(23576..23689,24080..24258,24347..24522,24635..24726)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translations="MTLGARLVAPMIILLFPIELVKAIEVITLTPTETSDKLANN
KAOLNGCISWCSFYSLSDTTRCLLSPCILLVQIEKDTAWPVKFCVPMCKHCKKLG
NWEDLKGAMEGDELEVEGDCGTSPACTVKEIHSYPTFMLFYNGEEVSKYKGRD
VESLKAFVVEETSKAAEKAQLEKEL"
complement(join(25423..25563,25675..26048,26127..26330,
26585..26783,26941..27842,27936..27991,28208..28413))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translations="MDSVRRKINNTTQSSSPKPSKFSVYRNPALAAASTANSIRP
SKSVLLFIFLLSTASAFSLTSFAGKEWLTNALTFGKISOEAYVTVKAWOGLVTLFC
IGAMMALSGISLHRAKFAARGTKSFSEKTDQPSLSNRQLLGIKKKADOSVSEF
PKSRPALQASLEPLVPIHQTLTGSAAKSSGGDKLNSRNGSLSSTPSPKQNGS
PSMYLVPSSPSVSSNLSSQDKVAPSWSGRRSSADITTEQLSOLLIADEKITE
SAGKMPPTPTVSGFAMASPTVSGTSGAGTRSTPLRVRNPSGAKFTPTPPKGE
GDGPNPMSLLETEGTHLGYPIEDWRDLRQWCSVLLKPLLNKVTSHIQVQGT
ASKWKLRTENQFOQOQOQOQOQALIPWQECVDALSEHRELQGLMKGEWYKGLLP
DASLQVLTQVTRALNEGTCVKNYEYNGAGDTRKNKWSLEPPDTHLLLLFLFCF
LEHPKWLHLDPSSYTGTOASKNPLFLGVLPPKRPFEKRYIAVVSVPSTLHPGCVL
AVDQSPPTFALYWDKRVQTLQGTALMDMLLICHRIKVGYGYYVGRGNLGGALL
ILQVDSDDTD"
complement(join(29418..29491,29579..29620,29858..29942,
30386..30805))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translations="WVSSKPKPKKARSDDVNVNKAASGRSKSSGSRTKTSNKVIVK
KKEIYEIISSESSDSVEEAIKGDEAKNGSVVSKRNGKSVGIPTKTSKNREDDGG

ASDAKIKPMNRIRIMRSDNSAPQIMQDAVFLVKNKATEMFIERPSEAYDSVSKDKK
KFIHYKHSVSVNDQRYEFLADSVPEKKAELAEEWERMGTDA"
Query Match 100.0%; Score 500; DB 8; Length 120965;
Best Local Similarity 100.0%; Pred. No. 1.3e-77;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATATTAAAGTTGGTTTATATCTTCTTAGTAATCTTTGAAATATATGTAAGAGATAATG 60
Db 67683 AATATTAAAGTTGGTTTATATCTTCTTAGTAATCTTTGAAATATATGTAAGAGATAATG 67624
QY 61 CTTCTAATAATAACATTCGATTTATTGGAATTAATGTAATGTAATGTAATGTAATGTAATGTAATG 120
Db 67623 CTTCTAATAATAACATTCGATTTATTGGAATTAATGTAATGTAATGTAATGTAATGTAATG 67564
QY 121 TACAGTGTATTTTGGAAACGACCAAAATGATATATGTAATGTAATGTAATGTAATGTAATGTAATG 180
Db 67563 TACAGTGTATTTTGGAAACGACCAAAATGATATATGTAATGTAATGTAATGTAATGTAATGTAATG 67504
QY 181 TAGTGTAAATAGGATACGGGACAAAGGTTGATCGACTCTAAACATATATGGGTACGTAATTC 240
Db 67503 TAGTGTAAATAGGATACGGGACAAAGGTTGATCGACTCTAAACATATATGGGTACGTAATTC 67444
QY 241 GCAGTGGTTACAGTCTACTGTGCGAGGCCAAACCTGTAATTAACGTTTGAAGTTTAGAGA 300
Db 67443 GCAGTGGTTACAGTCTACTGTGCGAGGCCAAACCTGTAATTAACGTTTGAAGTTTAGAGA 67384
QY 301 AATATTTTGTATGATAGTAGTACCAACAAATCAAGATGATAGTGTGTTTAATCAGTGTAAAAATGT 360
Db 67383 AATATTTTGTATGATAGTAGTACCAACAAATCAAGATGATAGTGTGTTTAATCAGTGTAAAAATGT 67324
QY 361 TGATCGAATCTACGAATGCAGAACATACATATATTTTAAATCTCTTTGGAATTTTGT 420
Db 67323 TGATCGAATCTACGAATGCAGAACATACATATATTTTAAATCTCTTTGGAATTTTGT 67264
QY 421 TTTGTTTTTATCATTTTTTGAATACACAGAGCTCAGTTATATTTTATATTTGATATGAA 480
Db 67263 TTTGTTTTTATCATTTTTTGAATACACAGAGCTCAGTTATATTTTATATTTGATATGAA 67204
QY 481 TTTGTTCTATTTTAAATCTTCA 500
Db 67203 TTTGTTCTATTTTAAATCTTCA 67184
RESULT 4
LOCUS HSA333293 696 bp DNA linear PRI 18-JUL-2002
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
NL6-EM8RS.
ACCESSION AJ333293
VERSION AJ333293.1 GI:15877711
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 696)
AUTHORS Kutsenko, A.S., Giszatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,
Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V.,
Levitsky, V.G., Kolchanov, N.A., Protogopov, A.I., Kaehuba, V.I.,
Kisselev, L.I., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.
TITLE NotI flanking sequence: a tool for gene discovery and verification
of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 696)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden Location/Qualifiers

```

source      1..696
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="NL6-EM8RS"

ORIGIN
Query Match      11.6%; Score 58; DB 9; Length 696;
Best Local Similarity 48.8%; Pred. No. 1.1;
Matches 244; Conservative 0; Mismatches 250; Indels 6; Gaps 3;

QY      2 ATATTTAAGTTTGGTTTATATCTTCTAGTAATCTTTGAAATATTGTAAGATATGCG 61
Db      136 ATAAATTAATAAATTTTATATTAATAATTAATTAATTTTATATTAATTAATTAATTAAT 195
QY      62 TTCTAATAAA--ATAACATTTGATTTATTTGGAATTAATGTAATTTGAAACAACTATGCAATA 119
Db      196 TTTTAATAATTAATTTTATTTTATTTTATTTTATTAATTAATTAATTAATTAATTTTATA 255
QY      120 CTACAGTGTATTTTGGACGACCAAAATGATATATGTAATGTAATCTTTGCTTCTTCTAC 179
Db      256 TTATTTATTTATTTTAAACATATAATAATATATATATATATATTTTATATTTTAAATTTAT 315
QY      180 ATAGTGAATAGGATAGCGACAGGTTGATCGACTCTA-AACATTTATGGTACGTAATT 238
Db      316 TTATTTATTTATTTTATAATAAATTTATATATAATTAATAATAATTAATTTATTTATTAATT 375
QY      239 CCGCAGTGGTTACAGTCTACTGTCTGCGGCCAACTGGTAAATTAACCGTTTGAAGTTTGA 298
Db      376 ATATATATGTTATATATATTTTATTTTATTA--TTTATATATTTTATATATATATA 432
QY      299 GAAATATTTTGATGATGAGTACCAACAATCAAGATGATAGTGTTATCACTGPAATAAT 358
Db      433 TATATTTAAATTTTATATTTTATATTAATTAATATATATATAAATTTTATTTAAATAATAAT 492
QY      359 GTTCAATTTGATACACGAATGACAGACATATACATATTTTAACTCTTTGGAAATTTTG 418
Db      493 TTTATTTAATATATATATTTAATTAATTTTAAATATATATTTATTTATATATAATTAATAA 552
QY      419 TTTTGTGTTTATCAATTTTGAATACACGAGAGCTCAGTTATATATTTTCATATTTGATATG 478
Db      553 ATATATTTATTTTCAATTTTATATTAATTAATTAATTAATTTCTTATTTCTTAAATTTATTA 612
QY      479 AATTGTTCTTATTAATCTT 498
Db      613 AATATGTTAATAATAATTATT 632

RESULT 5
BX957335      170606 bp      DNA      linear      HTG 03-APR-2004
LOCUS      Danio rerio clone CH211-119A10, WORKING DRAFT SEQUENCE, 9 unordered
DEFINITION      pieces.
ACCESSION      BX957335
VERSION      BX957335.3 GI:46200401
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 170606)
Burton,J.
Direct Submission
Submitted (02-APR-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Apr 5, 2004 this sequence version replaced gi:42821047.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk

----- Project Information
Center project name: zcl19A10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 168054 bases at least Q40
Consensus quality: 168519 bases at least Q30
Consensus quality: 168785 bases at least Q20
Insert size: 169806; sum-of-contigs
Insert size: 118142; 45.1% error; agarose-fp
Quality coverage: 6.01x in Q20 bases; sum-of-contigs Quality
coverage: 8.84x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2430: contig of 2430 bp in length
* 2431 2530: gap of 100 bp
* 2531 15089: contig of 12559 bp in length
* 15090 51042: contig of 35853 bp in length
* 51043 51142: gap of 100 bp
* 51143 54917: contig of 3775 bp in length
* 54918 55017: gap of 100 bp
* 55018 69969: contig of 14951 bp in length
* 69969 70069: gap of 100 bp
* 70069 103365: contig of 33297 bp in length
* 103366 103466: gap of 100 bp
* 103466 115110: contig of 11645 bp in length
* 115111 115210: gap of 100 bp
* 115211 167056: contig of 51846 bp in length
* 167057 170606: contig of 100 bp
* 167157 170606: contig of 3450 bp in length.
*
FEATURES
source
1..170606
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-119A10"
/clone_lib="CHORI-211"
1..2430
/note="assembly fragment:00021"
fragment_chain:1
clone_end:SP6
vector_side:left
2531..15089
/note="assembly fragment:00249"
fragment_chain:1
15190..51042
/note="assembly fragment:00571"
fragment_chain:1
51143..54917
/note="assembly fragment:00046"
fragment_chain:1
55018..69968
/note="assembly fragment:00397"
fragment_chain:1
70069..103365
/note="assembly fragment:00918"
fragment_chain:1
103466..115110
/note="assembly fragment:00117"
fragment_chain:2
115211..167056
/note="assembly fragment:01320"
fragment_chain:2
167157..170606
/note="assembly fragment:00081"

```



```
QY 25 TTCTAGTAATCTTTGAAATATTGTAAGAGATAATGCTTCTAATAATAACATTTGGATTT 84
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3904 TATTTATTTATATGATATATATATATAATAATAATAATAATAATAATAATAATAATA 3963
QY 85 ATTGGAATTAATGATTGAAAGAACTATGCAAACTACTACAGTGTATTTTGGACGACCAA 144
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3964 ATATATATATATATATATATATATATATATATATATATATATATATATAATAAATAAT 4023
QY 145 AATGATATATGTAACCTTTTCCTAGTCTTCTACATAGTGTAAATAGGATAGCGACAAG 204
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4024 ATTTATATATATATATATATATATATATATATATATATATATATATATATAATAAATAAT 4083
QY 205 GTTGATCGACTTAAACATTTATCGGTACGTAATTCGCCAGTGGTTACAGTCTACTGTCCA 264
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4084 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4143
QY 265 GCGCAAACTGGTAAATTAAAGCTTTGAGTTTACAGAAATATTTTGATGATGAGTACCACA 324
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4144 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4203
QY 325 ATCAAGATGATAGGTGTTAATCACTGTAATAATGTTGATGTAATCACTACGAATGCAGAA 384
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4204 ATNNATATATANTATTTTATTTTATTTTNNNAAATTTANNATATAAATAAANNTTA 4263
QY 385 CATATACATATTTTAACTCTTTTGGAAATTTTGTGTTTGTGTTTATCATTTTGTGAATAC 444
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4264 TTTTNTTAAATTTATTTNTTNTTNNATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 4323
QY 445 ACGAAGCTCAGTATATTTTCAATTTGATATGTAATTTGTTCTATTTAAT 495
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4324 AATAAAAAAANAANNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 4374
```

```
RESULT 8
LOCUS AY672413 527 bp DNA linear INV 31-JUL-2004
DEFINITION Meloidogyne partityla cytochrome oxidase subunit II (COII)
pseudogene, partial cds; tRNA-His gene, complete sequence; and 16S
ribosomal RNA gene, partial sequence; mitochondrial.
SOURCE AY672413 GI:50593125
ORGANISM mitochondrion Meloidogyne partityla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 527)
Thomas, S.H., Potenza, C., Jacobson, A.L. and Fuchs, J.M.
Differentiation of Meloidogyne partityla from other root-knot
nematode species in pecan using mt-DNA
Unpublished
REFERENCE 2 (bases 1 to 527)
Thomas, S.H., Potenza, C., Jacobson, A.L. and Fuchs, J.M.
Direct Submission
Submitted (29-JUN-2004) Entomology, Plant Pathology and Weed
Science, New Mexico State University, Skeen Hall, Room N141, MSC
3BE, Las Cruces, NM 88003, USA
```

```
FEATURES
source
1..527
/organism="Meloidogyne partityla"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:285957"
/country="USA: Cochise County, Arizona"
<1..93
/gene="COII"
/pseudo
<1..93
/gene="COII"
/notes="cytochrome oxidase subunit II"
/pseudo
/codon_start=1
/transl_table=5
```

```
tRNA 92..145
rRNA /product="tRNA-His"
146..>527
/product="16S ribosomal RNA"
ORIGIN
Query Match 11.3%; Score 56.4; DB 3; Length 527;
Best Local Similarity 47.0%; Pred. No. 2.2;
Matches 206; Conservative 0; Mismatches 231; Indels 1; Gaps 1;
QY 62 TTCTAATAATAACATTTGGAATTTATGGAATTAATGTAATTTGAAAAA-ACATATCAATATC 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 TTATACCTTATGATGATGTAATTTATTTATTTATTTTAAAAAATAATATTATTAGCT 97
QY 121 TACAGTGTATTTTGGAAACGACCAAAATGATATATATGTAAAACTTTTCGTTCTAGTCTTCTACA 180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 ATAAATTTCTAATGAATTTTATTTGATTAATAAAGTTTATTAGCTTTTATTATTAATGT 157
QY 181 TAGTGTAAATAGGATAGCGGACAGGTTGATCGACTCTAAACATTATGGGTACGTAATTC 240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 CATTAATTTTATTTGTTGAAAGAAATTAATAATTTGTTGTTTAAATAAAAAATTTAA 217
QY 241 GCAGTGGTTACAGTCTACTGTCGAGGCCAACTGGTAATTAACGTTTGAAGTTTAGAGA 300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 ATGATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 277
QY 301 AATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 AATTTTGTGTTAATAATATATATAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 337
QY 361 TGATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 397
QY 421 TTTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 480
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 457
QY 481 TTTGTTTCTATTTAATCTT 498
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
458 TTTAATTTTATTTT 475
```

```
RESULT 9
LOCUS CR627484 162190 bp DNA linear HTG 10-AUG-2004
DEFINITION Danio rerio clone CH211-281M7, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
ACCESSION CR627484
VERSION CR627484.3 GI:51100605
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 162190)
Burton, J.
Direct Submission
Submitted (09-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 10, 2004 this sequence version replaced gi:51036467.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC281M7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
```


Consensus quality: 18204 bases at least Q40
 Consensus quality: 18961 bases at least Q30
 Consensus quality: 13679 bases at least Q20
 Insert size: 161390; sum-of-configs
 Insert size: 160611; 1.1% error; agarose-
 Quality coverage: 10.48x in Q20 bases; sum-
 coverage: 10.53x in Q20 bases; agarose-
 coverage: 10.53x in Q20 bases; agarose-

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

2406:	contig of 2406 bp in length
2506:	gap of 100 bp
52017:	contig of 49511 bp in length
52117:	gap of 100 bp
52118:	contig of 13844 bp in length
65961:	contig of 100 bp
66061:	gap of 100 bp
80113:	contig of 14052 bp in length
80213:	gap of 100 bp
84677:	contig of 4464 bp in length
84777:	gap of 100 bp
104173:	contig of 19396 bp in length
104273:	gap of 100 bp
142032:	contig of 37759 bp in length
142132:	gap of 100 bp
159338:	contig of 17206 bp in length
159438:	gap of 100 bp
162130:	contig of 2752 bp in length.

FEATURES

source

```

location/Qualifiers
1. .162190.
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-281M7"
/clone_lib="CHORI-211"

```

misc feature

```

1. .2406
/note="assembly_fragment:00045
fragment chain:1

```

```
clone_end:SP6
vector_side:left"
2507.52017
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

/note="assembly fragment:01809
fragment_chain:1"
52118. .65961

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

/note="assembly_fragment:00327
fragment_chain:1"
66062.80113

```

1

```

/note="assembly_fragment:00139
fragment_chain:2"
80214. .84677

```

1

```
/note="assembly_fragment:00083
fragment_chain:2"
84778. .104173
```

1

```
/note="assembly fragment:00821
fragment_chain:2"
104274. 142032
```

1

```

/note="assembly_fragment:01182
fragment_chain:2"
142133. .159338

```

```

/note="assembly_fragment:00554
fragment_chain:2"
159439. .162190

```

```
/note="assembly fragment:00012
fragment_chain:2
clone_end:T7
```

ORIGIN

Query Match	11.2%;	Score 56.2;	DB 2;	Length 162190;
Best Local Similarity	45.9%;	Pred. No. 0.9;		
Matches 228; Conservative	0;	Mismatches 268;	Indels 1;	Gaps 1;
<hr/>				
Qy	2	ATATTTAAAGTTGGTATTATTCCTTCTAGTAATCTTTGCAATATGTGAAGATAATGCC	61	
Db	41340	ATATTATATTATATATATATTATATATTAATATATATATATATATATATATATATTA	41399	
<hr/>				
Qy	62	TTCCTAATAACACTGGATTATTTGGAAATTAATGTATGAAAAAACTATGCAAAATACT	121	
Db	41400	TTATATATTATATTAATTAATTTATATATATATATATATATATATATATATATATA	41459	
<hr/>				
Qy	122	ACAGTGATTTTTGGAACGACCAAAANGATATATATATATATATATATATATATATAT	181	
Db	41460	TTAT	41519	
<hr/>				
Qy	182	AGTGTAATAGGATAGCGGCACAGGTGATCGACTCAAACATATATGGGTACGTAAATCCG	241	
Db	41520	TAT	41579	
<hr/>				
Qy	242	CAGTGGTTACAGTCTACTCTCAGGCGCAAACTGGGTAAATTAACGTTTGAAGTTTAGAGA	301	
Db	41580	AATTAT	41639	
<hr/>				
Qy	302	ATATTTTGAATGATGATCCACAATCAAGATGATAGGTGTTAATCACTCTGAAAAATGTT	361	
Db	41640	TTAAAT	41699	
<hr/>				
Qy	362	GATTGAACTACTAGGAATGCAGAACATATACATATTTTAAATCTCTTTGGAAATTTTGT	421	
Db	41700	TATTAT	41759	
<hr/>				
Qy	422	TTGTTTTTATCATTTTTTGGAATACACGAAGAGCTCAGTTATATATTTTCATATTCGAT	481	
Db	41760	AATTAT	41818	
<hr/>				
Qy	482	TTGTCTCATTTAAATCTT	498	
Db	41819	TTATATATATATATATAT	41835	

RESULT 10
BX088594/c

LOCUS
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

ORGANISM

REFERENCE

AUTHORS

AUTHORS	TITLE	JOURNAL
W. J. G. Brouwer J. A. M. M. van den Broek A. H. M. M. van der Meulen H. J. A. M. M. van der Meulen	THE EFFECT OF THE CUTTING PLANE ON THE MECHANICAL PROPERTIES OF THE CUTTING PLANE	Journal of Applied Polymer Science

JOURNAL

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml RP71-1G17 is from a Zebrafish mixed sex BAC library VECTOR: pTARBAC2.

FEATURES

source
1. .144022
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="RP71-1G17"
/clone_lib="RPCI-71"

ORIGIN

Query Match 11.1%; Score 55.6; DB 5; Length 144022;
Best Local Similarity 50.4%; Pred. No. 1.2;
Matches 136; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
QY 223 TTATGGGTACGTAATCCGAGTGGTTACAGTCTACTGCGAGGCCAACTGGTATTAA 282
DB 98932 TTATGTGAACGCTGCTGTTCTTTTCAACCAATTTGTTATTAATTAATTTTCATAAT 98873
QY 283 ACCTTTGAATTTAGAGAATAATTTTGATGATCAGTACCACATCAAGATGATGCTGT 342
DB 98872 TTATTTTACAAATTTTAAACATTTTAAACATCAAGAAATCAATATATTTTAT 98813
QY 343 TAATCACTGAAATGTTGATGTAATGTAATGCAAGTGCAGAACATATACATATTTTAAAT 402
DB 98812 TTATTTATGTAATATATAGAGCATTTATTTATTTATTTATTTATTAATATTTATTT 98753
QY 403 CTCTTTGGAATTTTCTTTTGTGTTTATCATTTTGAATACAGAGAGCTCAGTTATA 462
DB 98752 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 98693
QY 463 TTTCATATTTGATATGAAATTTGTTCTATTT 492
DB 98692 TTTTGTGAATGTTAATTTTCTTTTGTAAAT 98663

RESULT 11

CR788237/c 228431 bp DNA linear HTG 23-SEP-2004
LOCUS Danio rerio clone DKEY-75L11, *** SEQUENCING IN PROGRESS ***, 16
DEFINITION unordered pieces.

ACCESSION CR788237
VERSION CR788237.4 GI:52626720
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 228431)
AUTHORS McLay, K.

TITLE Direct Submission

JOURNAL

Submitted (22-SEP-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 23, 2004 this sequence version replaced gi:52353229.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zf1sh-help@sanger.ac.uk

----- Project Information

Center project name: zk75L11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 223308 bases at least Q40

Consensus quality: 224622 bases at least Q30

Consensus quality: 225480 bases at least Q20

Insert size: 226931; sum-of-contigs

Insert size: 231664; 2.2% error; agarose-fp

Quality coverage: 8.43x in Q20 bases; sum-of-contigs Quality

coverage: 8.33x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2629: contig of 2629 bp in length
* 2630 2729: gap of 100 bp
* 2730 19162: contig of 16433 bp in length
* 19163 19262: gap of 100 bp
* 19263 49824: contig of 30562 bp in length
* 49825 49924: gap of 100 bp
* 49925 56858: contig of 6934 bp in length
* 56859 56959: gap of 100 bp
* 56959 66911: contig of 9952 bp in length
* 66911 67010: gap of 100 bp
* 67011 69530: contig of 2520 bp in length
* 69531 69631: gap of 100 bp
* 69631 105513: contig of 35883 bp in length
* 105514 105614: gap of 100 bp
* 105614 109364: contig of 3751 bp in length
* 109365 109464: gap of 100 bp
* 109465 113382: contig of 3918 bp in length
* 113383 113482: gap of 100 bp
* 113483 120793: contig of 7311 bp in length
* 120794 120893: gap of 100 bp
* 120894 136428: contig of 15535 bp in length
* 136429 136528: gap of 100 bp
* 136529 138587: contig of 2059 bp in length
* 138588 138687: gap of 100 bp
* 138688 153515: contig of 14828 bp in length
* 153516 153616: gap of 100 bp
* 153616 175666: contig of 22051 bp in length
* 175667 175767: gap of 100 bp
* 175767 200608: contig of 24842 bp in length
* 200609 228431: contig of 100 bp
* 228431 228431: contig of 27723 bp in length.

FEATURES

source

1. .228431
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-75L11"
/clone_lib="DanioKey"

misc_feature

1. .2629
/note="assembly_fragment:00084
fragment chain:1"
2730. .19162

misc_feature

/note="assembly_fragment:01618"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 381.991 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929c-1_COPY_1_500

Perfect score: 500
Sequence: 1 aatattagtttggttat.....ttgtctatttaatttca 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10329	4 AAA89385	Aaa89385 Arabidops
2	500	100.0	10329	4 AAA89353	Aaa89353 Arabidops
3	139	27.8	1085	3 AAC36438	Aac36438 Arabidops
4	54.6	10.9	50000	6 ABL56201	AbL56201 AmbPV gen
5	52.8	10.6	802	13 ADR64347	Adr64347 Cotton cD
6	52.6	10.5	14548	6 ABL70603	AbL70603 Chemicall
7	51	10.2	13919	6 ABL92305	AbL92305 Chemicall
8	51	10.2	158001	12 ADL17884	AdL17884 Human pho
9	50.8	10.2	8246	6 ABL32202	AbL32202 Human imm
10	50.6	10.1	7072	6 ABK31471	Abk31471 Signal tr
11	50.6	10.1	7072	6 ABL70566	AbL70566 Chemicall
12	50.6	10.1	7072	6 AAS61385	Aas61385 Human gen
13	50.4	10.1	5912	6 ABL32602	AbL32602 Human imm
14	50.2	10.0	8056	8 ABZ10246	Abz10246 Haematopo
15	49.8	10.0	15832	4 AAS45399	Aas45399 Chemicall
16	49.8	10.0	15832	6 ABL33343	AbL33343 Human imm
17	49.8	10.0	15832	6 ABK28244	Abk28244 DNA trans
18	49.8	10.0	34688	6 ABQ67059	Abq67059 Human ang
19	49.6	9.9	732	2 ADR02344	Adr02344 A. gossyp
20	49.6	9.9	5379	6 ABL33676	AbL33676 Human imm

21	49.6	9.9	5379	6 ABL34576	AbL34576 Human met
22	49.6	9.9	5379	7 ADS99837	AdS99837 Chemicall
23	49.6	9.9	5379	7 ADS99837	AdS99837 Bisulphit
24	49.6	9.9	6880	6 ABK31323	Abk31323 Signal tr
25	49.6	9.9	6880	6 ABL70294	AbL70294 Chemicall
26	49.6	9.9	6880	6 AAS61224	Aas61224 Human gen
27	49.6	9.9	8056	8 ABZ10246	Abz10246 Haematopo
28	49.4	9.9	346	4 AA187279	AaI87279 Human pol
29	49.2	9.8	17131	6 ABL33052	AbL33052 Human imm
30	48.6	9.7	5379	6 ABL33677	AbL33677 Human imm
31	48.6	9.7	5379	6 ABL34577	AbL34577 Human met
32	48.6	9.7	5379	6 ABL70370	AbL70370 Chemicall
33	48.6	9.7	5379	7 ADS99838	AdS99838 Complemen
34	48.6	9.7	5567	6 ABL33589	AbL33589 Human imm
35	48.6	9.7	5567	6 ABL92273	AbL92273 Chemicall
36	48.2	9.6	5520	6 ABL33519	AbL33519 Human imm
37	47.8	9.6	34688	6 ABQ67060	Abq67060 Human ang
38	47.6	9.5	5464	6 ABL33917	AbL33917 Human imm
39	47.6	9.5	5464	6 ABK31453	Abk31453 Signal tr
40	47.6	9.5	5464	6 ABL70430	AbL70430 Chemicall
41	47.6	9.5	5546	6 ABL32391	AbL32391 Human imm
42	47.6	9.5	5546	6 ABK34005	Abk34005 Human DNA
43	47	9.4	8056	8 ABZ10100	Abz10100 Haematopo
44	47	9.4	8237	4 AAS46802	Aas46802 Tumour su
45	47	9.4	9721	4 ABL07586	AbL07586 Drosophil

ALIGNMENTS

RESULT 1

AAA89385

ID AAA89385 standard; DNA; 10329 BP.

XX AAA89385;

XX 04-NOV-2004 (revised)

DT 23-APR-2001 (first entry)

XX Arabidopsis thaliana gene involved in epigenetic gene silencing.

XX Gene silencing; silencing gene; MOM, ds.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT exon 1..1008

FT /tag= a

FT /number= 1

FT intron 1009..1295

FT /tag= b

FT /number= 1

FT exon 1296..2550

FT /tag= c

FT /number= 2

FT CDS 1301..9808

FT /tag= d

FT /product= "Columbia protein"

FT /note= "this sequence contains 15 introns."

FT intron 2551..2673

FT /tag= e

FT /number= 2

FT exon 2674..2752

FT /tag= f

FT /number= 3

FT intron 2753..2867

FT /tag= g

FT /number= 3

FT exon 2868..3113

FT /tag= h

FT /number= 4

FT intron 3114..3506

FT /tag= i

XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX Claim 6; Page 19-22; 48pp; English.
XX The present sequence is that of an Arabidopsis thaliana gene (MOM)
CC involved in epigenetic gene silencing. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia, occurring at nucleotide positions
CC 4338 (A instead of T) and 6721 (T instead of G). Gene silencing is useful
CC as a molecular tool for regulating gene expression
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
SQ Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;
Query Match 100.0%; Score 500; DB 4; Length 10329;
Best Local Similarity 100.0%; Pred. No. 1.2e-86;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATATTAAAGTTGGTTTATTTCTTCTAGTAACTCTTGAATATTTGTAAGAGATAATG 60
DB 1 AATATTAAAGTTGGTTTATTTCTTCTAGTAACTCTTGAATATTTGTAAGAGATAATG 60
QY 61 CTTCTAATAAATCAATTTGGATTATTTGGAAATTAATGTAATGTAAGAAACTATGCAATAC 120
DB 61 CTTCTAATAAATCAATTTGGATTATTTGGAAATTAATGTAATGTAAGAAACTATGCAATAC 120
QY 121 TACAGTGTATTTTGGACGACCAAAATGATATATGTAATCTTCGTTCTAGTCTTCTACA 180
DB 121 TACAGTGTATTTTGGACGACCAAAATGATATATGTAATCTTCGTTCTAGTCTTCTACA 180
QY 181 TAGTGTAAATAGGATAGCGGACAGGTTGATCGACTCTAAACATTTATGGGTACGTAATTC 240
DB 181 TAGTGTAAATAGGATAGCGGACAGGTTGATCGACTCTAAACATTTATGGGTACGTAATTC 240
QY 241 GCAGTGGTTACAGTCTACTCTCGAGGCCAACTGGTAATTAATGTAAGTTGAGTTAGAGA 300
DB 241 GCAGTGGTTACAGTCTACTCTCGAGGCCAACTGGTAATTAATGTAAGTTGAGTTAGAGA 300
QY 301 AATATTTTGATGATGAGTACCAATCAAGATGATAGGTTAATCACTGTAAAAATGT 360
DB 301 AATATTTTGATGATGAGTACCAATCAAGATGATAGGTTAATCACTGTAAAAATGT 360
QY 361 TGATTGAATACACTAGAAATGAGAACATATATATTTTAACTCTCTTTGGAATTTTGT 420
DB 361 TGATTGAATACACTAGAAATGAGAACATATATATTTTAACTCTCTTTGGAATTTTGT 420
QY 421 TTTGTTTTTATCATTTTTTGGATACAGGAGCTCAGTTATTTTTCATATTTGATATGAA 480
DB 421 TTTGTTTTTATCATTTTTTGGATACAGGAGCTCAGTTATTTTTCATATTTGATATGAA 480

QY 481 TTTGTTCTATTAACTTCA 500
DB 481 TTTGTTCTATTAACTTCA 500
RESULT 3
AAC36438
ID AAC36438 standard; DNA; 1085 BP.
XX AAC36438;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 13805.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-014091P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 27.8%; Score 139; DB 3; Length 1085;

Best Local Similarity 100.0%; Pred. No. 1.4e-17; Mismatches 0; Indels 0; Gaps 0;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AATATTTAAGTTGGTTTATTTCTTTCTAGTAATCTTTGAAATATTTGTAAGATAATG 60

Db 947 AATATTTAAGTTGGTTTATTTCTTTCTAGTAATCTTTGAAATATTTGTAAGATAATG 1006

Qy 61 CTTCTAATAAATAACATTTGGAATTTATTTGGAATTAATGTTATGAAAAAATCTGCAATAC 120

PF 29-JUN-2001; 2001WO-EP007472.
XX
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-147896/19.
DR
XX
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
XX Claim 1; SEQ ID NO 314; 24pp; English.
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
XX Sequence 7072 BP; 2088 A; 51 C; 1418 G; 3515 T; 0 U; 0 Other;
SQ
Query Match 10.1%; Score 50.6; DB 6; Length 7072;
Best Local Similarity 52.1%; Pred. No. 1.3;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 279 TTAACGTTTGAAGTTTAGAGAAATATTTTGATGATGAGTACCACCAATCAAGATGATAG 338
DB 6361 TTGAATTTTGTAGTTTAAAGTTATTTATTTGTTATAGTTTAAAGTTGTGAGATTATAG 6420
QY 339 GTGTTAATCACTGTAAAAATGTTGATGTAATCTACGAATGCAGAACATATATATTTT 398
DB 6421 GTGTGAGTTATTATATTTGGTTTGTATATATTTTATAGATTTTATGATTTATTTTAT 6480
QY 399 TAATCTCTTTGGAAATTTTGTGTTTATCATTTTGTATATGATGATGATGATGATGATG 458
DB 6481 GAATTAATTTTAAATTTTGTGTTTATATTTTGTGTTTATTTTATGATTTTATGAT 6540
QY 459 TATATTTCATATGATATGATGAAATTTGTTCTATTTAAT 495
DB 6541 TTTATTAAATTTTGTGTTTATATTTTGTGTTTAT 6577
RESULT 11
ABL70566
ID ABL70566 standard; DNA; 7072 BP.
XX
AC ABL70566;
XX
XX 01-JUL-2002 (first entry)
XX Chemically treated cell signalling DNA sequence complementary to#228.
XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX

OS Unidentified.
XX
XX WO200202807-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007471.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154758/20.
XX
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
XX Claim 1; SEQ ID NO 456; 24pp + Sequence Listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
XX Sequence 7072 BP; 2088 A; 51 C; 1418 G; 3515 T; 0 U; 0 Other;
SQ
Query Match 10.1%; Score 50.6; DB 6; Length 7072;
Best Local Similarity 52.1%; Pred. No. 1.3;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 279 TTAACGTTTGAAGTTTAGAGAAATATTTTGATGATGAGTACCACCAATCAAGATGATAG 338
DB 6361 TTGAATTTTGTAGTTTAAAGTTATTTATTTGTTATAGTTTAAAGTTGTGAGATTATAG 6420
QY 339 GTGTTAATCACTGTAAAAATGTTGATGTAATCTACGAATGCAGAACATATATATTTT 398
DB 6421 GTGTGAGTTATTATATTTGGTTTGTATATATTTTATAGATTTTATGATTTATTTTAT 6480
QY 399 TAATCTCTTTGGAAATTTTGTGTTTATCATTTTGTATATGATGATGATGATGATGATG 458
DB 6481 GAATTAATTTTAAATTTTGTGTTTATATTTTGTGTTTATTTTATGATTTTATGAT 6540
QY 459 TATATTTCATATGATATGATGAAATTTGTTCTATTTAAT 495
DB 6541 TTTATTAAATTTTGTGTTTATATTTTGTGTTTAT 6577
RESULT 12
AAS61385
ID AAS61385 standard; DNA; 7072 BP.
XX
AC AAS61385;
XX
XX 29-JAN-2002 (first entry)
XX Human gene regulation-associated gene oligonucleotide #340.
DE
XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX
 OS Homo sapiens.
 PN WO200177375-A2.
 PD 18-OCT-2001.
 XX
 XX 06-APR-2001; 2001WO-EP003968.
 XX
 XX 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-017470/02.
 DR
 XX
 XX New nucleic acid sequences from chemically modified genes associated with
 PT gene regulation, useful for analyzing cytosine methylations for diagnosis
 PT and therapy of diseases e.g. severe combined immunodeficiency disease.
 XX
 XX Disclosure; SEQ ID NO 348; 26pp; English.
 PS
 XX The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The chemical
 CC pretreatment converts cytosine bases unmethylated at the 5-position to
 CC uracil or another base with hybridisation behaviour dissimilar to
 CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
 CC oligomers (or sets/arrays) and method are useful in the diagnosis of
 CC diseases (or predisposition to diseases) associated with gene regulation
 CC and in therapy of such diseases, by enabling analysis of the cytosine
 CC methylation patterns of such genes, kits are provided. They are
 CC especially useful in diagnosis and therapy of e.g. severe combined
 CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
 CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
 CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
 CC present sequence is a sequence included in the sequence data for this
 CC specification and is associated with the human gene regulation-associated
 CC genes. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7072 BP; 2088 A; 51 C; 1418 G; 3515 T; 0 U; 0 Other;
 SQ
 Query Match 10.1%; Score 50.6; DB 6; Length 7072;
 Best Local Similarity 52.1%; Pred. No. 1.3;
 Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
 QY 279 TTAACGTTTGAAGTTTGAAGAAATATTTTGATGATGAGTACCAATCAAGATGATAG 338
 DB 6361 TTGAATTTTGATTTAAGTTATTTATTTGTTATAGTTTAAAGTTTGAGATTATAG 6420
 QY 339 GTGTTATCACTGTAATAATGTTGATTAATCACTACCAATGCAGACATACATATTTT 398
 DB 6421 GTGTCGATTATATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 6480
 QY 399 TAATCTCTTGGAAATTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 458
 DB 6481 GAATATTTTAAATTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 6540
 QY 459 TATATTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495

Db 6541 TTTATTAATTTTGTGTTTATATATTTTGTGTTTAT 6577
 RESULT 13
 ABL32602
 ID ABL32602 standard; DNA; 5912 BP.
 XX ABL32602;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Human immune system associated gene SEQ ID NO: 575.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; neotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT
 XX Claim 1; SEQ ID NO 575; 32pp + Sequence Listing; German.
 PS
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX Sequence 5912 BP; 1670 A; 31 C; 1055 G; 3156 T; 0 U; 0 Other;
 SQ
 Query Match 10.1%; Score 50.4; DB 6; Length 5912;
 Best Local Similarity 52.4%; Pred. No. 1.4;
 Matches 111; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
 QY 287 TTGAAGTTTGAAGAAATATTTTGATGATGATGATGATGATGATGATGATGATGATGAT 346
 DB 2981 TTGAGAGATAGATATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 3040
 QY 347 CACTGTAAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
 DB 3041 TTGTTTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3100
 QY 407 TTGGAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 466
 DB 3101 GTAGTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 3160
 QY 467 ATATTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498


```
XX Claim 1; SEQ ID NO 104; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 15832 BP; 4164 A; 297 C; 3742 G; 7610 T; 0 U; 19 Other;
Query Match 10.0%; Score 49.8; DB 4; Length 15832;
Best Local Similarity 60.9%; Pred. No. 1.9;
Matches 81; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 360 TTGATTGAATACGAAATCGAGACATATACATATTTTAACTCTTTGGAATTTTGT 419
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7041 TTTTATTATATTAGAAATTATAAATTTGTAGATATTGTTAAATTTTGTGTTTTTGT 7100
QY 420 TTTGTTTTTATCATTTTTCAGATACAGAGAGCTCAGTTATATTTTCATATTGTATATGA 479
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7101 TTTTATAGAAATTTTTCGAAAATAAATTAATTTAGTTAAATTTAATTAATTTAGTTTT 7160
QY 480 ATTGTTCTATTT 492
Db || ||| |||
7161 TTTATATATATAT 7173
Search completed: July 12, 2005, 01:45:48
Job time : 387.991 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:04:38 ; Search time 114.348 Seconds
(without alignments)
7154.840 Million cell updates/sec

Title: US-10-018-929C-1_COPY_1_500
Perfect score: 500
Sequence: 1 aaatttaagtttggttat.....tttgtctatttaatttca 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	54.4	10.9	601	4	US-09-949-016-179398
C 2	53.2	10.6	263693	4	US-09-949-016-12386
C 3	53.2	10.6	263694	4	US-09-949-016-16915
C 4	49.6	9.9	732	3	US-08-998-416-1036
C 5	46.8	9.4	1150	1	US-08-161-406-1
C 6	46.6	9.3	854	3	US-08-998-416-534
C 7	46.5	9.3	860	3	US-08-998-416-287
C 8	46.2	9.2	19438	4	US-09-949-016-12699
C 9	46	9.2	19124	2	US-08-487-826B-13
C 10	45.8	9.2	663	3	US-08-998-416-191
C 11	45.2	9.0	601	4	US-09-949-016-179399
C 12	45.2	9.0	1141	4	US-09-806-708B-22
C 13	45	9.0	11474	4	US-09-495-406-1
C 14	45	9.0	11474	4	US-09-816-028A-1
C 15	45	9.0	11474	4	US-10-303-162-1
C 16	45	9.0	11474	4	US-10-303-134-1
C 17	44	8.8	820	4	US-09-270-767-29619
C 18	44	8.8	2389	4	US-09-270-767-13612
C 19	43.6	8.7	660	1	US-07-991-867B-32
C 20	43.6	8.7	660	1	US-08-107-755A-32
C 21	43.6	8.7	660	2	US-08-544-332-32
C 22	43.6	8.7	660	3	US-09-370-861A-32
C 23	43.6	8.7	1511	1	US-07-991-867B-8
C 24	43.6	8.7	1511	1	US-08-107-755A-8
C 25	43.6	8.7	1511	2	US-08-544-332-8
C 26	43.6	8.7	1511	3	US-09-370-861A-8
C 27	43.6	8.7	4810	3	US-08-852-629-11

C 28	43.6	8.7	4838	3	US-08-852-629-15	Sequence 15, Appl
C 29	43.6	8.7	6265	3	US-09-129-112-3	Sequence 3, Appl
C 30	43.6	8.7	18773	4	US-09-949-016-14164	Sequence 14164, A
C 31	43.4	8.7	12793	3	US-09-004-838-124	Sequence 124, App
C 32	43	8.6	724	3	US-08-998-416-683	Sequence 683, App
C 33	43	8.6	44353	4	US-09-949-016-15302	Sequence 15302, A
C 34	41.8	8.4	12730	3	US-09-004-838-91	Sequence 91, Appl
C 35	41.8	8.4	15062	3	US-09-004-838-89	Sequence 89, Appl
C 36	41.8	8.4	67755	4	US-09-949-016-13703	Sequence 13703, A
C 37	41.4	8.3	601	4	US-09-949-016-82196	Sequence 82196, A
C 38	41.4	8.3	601	4	US-09-949-016-189788	Sequence 189788, A
C 39	41.4	8.3	19167	4	US-09-949-016-17174	Sequence 17174, A
C 40	41.4	8.3	29037	4	US-09-949-016-14131	Sequence 14131, A
C 41	41.4	8.3	36242	4	US-09-949-016-12996	Sequence 12996, A
C 42	41.4	8.3	36242	4	US-09-949-016-12997	Sequence 12997, A
C 43	41.4	8.3	36242	4	US-09-949-016-12998	Sequence 12998, A
C 44	41.4	8.3	36242	4	US-09-949-016-12999	Sequence 12999, A
C 45	41.4	8.3	36242	4	US-09-949-016-13000	Sequence 13000, A

ALIGNMENTS

RESULT 1
US-09-949-016-179398/c
; Sequence 179398, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179398
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-179398

Query Match	10.9%	Score 54.4;	DB 4;	Length 601;
Best Local Similarity	46.0%	Pred. No. 0.0028;		
Matches 219;	Conservative 1;	Mismatches 252;	Indels 4;	Gaps 1;
Qy	1	AAATATTAAAGTTGGTTTATATCTTCTAGTAGTAATCTTCGAAATATTTGTAAGAGATATG 60		
Db	540	ATTATTTTATATATATATATATTTTATATATATATATATATATATATATATATATATATA 481		
Qy	61	CTTCTAATAATAACATTTGGAAATTAATGTAATTTGAAAAAATCTATGCAATATC 120		
Db	480	TATAATA 421		
Qy	121	TACAGTGATATTTGGAAACGACCAAAATGATATATATGTAACCTTCGTTCTAGTCTTCTACA 180		
Db	420	TAAAGTAAT 361		
Qy	181	TACTGTATAGGATAGCGGACAGGTGATCGACTCTAAACATTATGGGTACGTAATTC 240		
Db	360	ATATAAATGCAATACAT 301		
Qy	241	GCAGTGGTTACAGTCTACTGCTCGAGGCCAACTGGTAAATTAACAGTTTGAAGTTTAGAGA 300		
Db	300	TTTATATATAAAGTAATACTATATTTTATATACATAAAGTAATATATATATATATATA 241		
Qy	301	AATATTTTGATGATGAGTACCAACATCAAGATGATAGTGTGAATCACTGTAAAAATGT 360		

QY 410 GAATTTTGTGTTTGTGTTTATCATTTTGAATA 443
|||||
Db 1029 TAATTTCTGTTTATGTTTATTTTATAATA 996

RESULT 6

US-08-998-416-534
; Sequence 534, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgin
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; INFORMATION FOR SEQ ID NO: 534:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1372UP
; US-08-998-416-534

Query Match 9.1%; Score 46.6; DB 3; Length 854;
Best Local Similarity 44.2%; Pred. No. 0.16; 244; Indels 0; Gaps 0;
Matches 193; Conservative 0; Mismatches 244; Indels 0; Gaps 0;
QY 62 TTCTAATAATAACATTGGATTTTATGGAATTAATGATATGTAACCTTCGTTCTAGCTCTTACAT 121
Db 89 TACCATTTAATAATGATATATTATTCACAAAATTTATGTTAATAAATAAATAAATA 148
QY 122 ACAGTGTATTTTGGACGACCAAAATGATATATGTAACCTTCGTTCTAGCTCTTACAT 181
Db 149 ATTACTAATTAATAAATTTAATAATAAATAAAGATTTTAATTAATTTATATACCT 208
QY 182 AGTGTAATAGTAGCGGACAGGTTGATCGACTCTAAACATTTATGGTACGTAATTCG 241
Db 209 TAGATAATAAATTTAGATTTTAAAGTATCTTAAATAATATATATTATAGGTAATAATTTAG 268

QY 242 CAGTGTGTACGCTACTGTCGAGGCCAACTGGTAATTAACAGTTTGAAGTTTGAAGAA 301
|||||
Db 269 TAGGTAGTAATATCCAAATTAAGGGTAGACTATTAAATAGAAATATTACTAGACTAATA 328
QY 302 ATATTTTGTATGATGAGTACCAATCAAGATGATAGGTGTTAATCACTGCTAAATGTT 361
|||||
Db 329 AATAAATAATATTGAAAGGTACATTTAATAATATATATATATCAATCAAGTAAATTAATA 388
QY 362 GATTGAATACTACGAATGCAGAACATATACATATTTTAAATCTCTTTGGAATTTTGT 421
|||||
Db 389 ATTATATAAATAATTAATATATATACATTAATAATTAATAACTTAATATCTATTATTA 448
QY 422 TTGTTTTATCATTTTGTGAATACAGGAGCTCAGTTATATTTTCATATTTGTATATGAAT 481
Db 449 ATAAAAATGGTATATTTTAATTTAAATTTAAATTAATAACTATATTAAATAATTTCTAAA 508
QY 482 TTGTTCTATTATCTT 498
Db 509 GTAATTTCTATTATTATT 525

RESULT 7

US-08-998-416-287
; Sequence 287, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgin
; APPLICANT: Knechtel, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; INFORMATION FOR SEQ ID NO: 287:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 860 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1240UP
; US-08-998-416-287

䷗

; FILE REFERENCE: 4810-58741

```
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEl promoters
US-09-806-708B-22

Query Match          9.0%; Score 45.2; DB 4; Length 1141;
Best Local Similarity 9.9%; Pred. No. 0.35;
Matches 46; Conservative 194; Mismatches 223; Indels 0; Gaps 0;

QY 38 TTGAAATATTGTAAGAGATATGCTTCTAATAATAACATTGGATTATTGGAAATTAATG 97
Db 629 TMRWTKMGDGTVRKVKWEDTCTTYDVWADSWVWYANWRCRDVYTRNNYCKSYA 570
QY 98 TATTGAAAAAATGACAAATACATACAGTGTATTTTGGAAACGACCAAAATGATATGTA 157
Db 569 HSYWYWSNNAMVRRYSARNWSSWARTTRNNWMSGBVRWAGTMMWRHNNNNNDTR 510
QY 158 AACTTCGTTCTAGTCTTCTACATAGTGTATAGGATAGCGGACAGGTTGATCGACTCT 217
Db 509 YYWKKRWABRTTVYDSMCAKSMWRGNWRAKMMWMAANNDAGAMDHWTYWMGNNTMW 450
QY 218 AAACATTATGGGTACGTAATTCGCGAGTGTCTACAGTCTACTGTGAGGCCAAACTGGTA 277
Db 449 MRRANKMNMWACERAYCCNNNNNRACVHKHKMRWTWYKMKACNNNNBKMYRVA 390
QY 278 ATTAACAGTTTGAAGTTTAGAGAAATATTGATGATGATGATACACAAATCAAGATGATA 337
Db 389 WMYSRDVTNTDMMWTSDBWBYWYVDYTMRAWNNNNNNNRBCKTTSWMMWMDHMT 330
QY 338 GGTGTTAATCACTGTAAATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 397
Db 329 HCTYGNNTWSAYBMAWSMAWAGASNBVTYNCWRWYMGKTMNNNNNKAWYRTKT 270
QY 398 TTAATCTCTTTGGAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 457
Db 269 VANCNNRYDYTAWTEKRYKYCYABWYBYBYMYMGKHBBWRRAHRSWNNWVKCRN 210
QY 458 TTATATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
Db 209 KYMVSWHYHAMRYBKWABAVGCNNWKKDRMAHHHCATNNNMW 167

RESULT 13
US-09-495-406-1/c
; Sequence 1, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-816-028A-1
```

```
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

Query Match          9.0%; Score 45; DB 4; Length 11474;
Best Local Similarity 49.0%; Pred. No. 0.56;
Matches 120; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 252 AGTCTACTGCGAGGCCAACTGCTAATAAAGCTTTGAAGTTTAGAGAAATATTTTGAT 311
Db 5211 AGTATTGTAGCGTTCTCTATGCTAATAAATTTTGTCTATATATTCAAAATATATAAAA 5152
QY 312 GATGAGTACCACAAATCAAGATGATAGGTGTTTAATCACTGTAAAAATGTTGATTGAATAC 371
Db 5151 TAAAAAGACAAAATTAATGGTTACATTATTAGCTAAATAAATATTTTAACITTTAAAAA 5092
QY 372 TAGCAATGCAGAACATATACATATTTTAACTCTCTTTGGAAATTTTGTGTTTTGTTTAT 431
Db 5091 TTTTATTAAATATAAAAAATTTTAAAAACAATGTTTAAAGAAATATTTTTTTAGAGTTTCT 5032
QY 432 CATTTTGTATACACGAGAGCTCAGTTATATTTTATATTTGATATCAATTTGTTCTTATT 491
Db 5031 TTTGAAATCATAGAGGTTCTATTATTTATGATTAAATATTTTGTAAATAGAAATTTTAAAT 4972
QY 492 TAAATC 496
Db 4971 TCTTC 4967

RESULT 14
US-09-816-028A-1/c
; Sequence 1, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-816-028A-1

Query Match          9.0%; Score 45; DB 4; Length 11474;
Best Local Similarity 49.0%; Pred. No. 0.56;
Matches 120; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 252 AGTCTACTGCGAGGCCAACTGCTAATAAAGCTTTGAAGTTTAGAGAAATATTTTGAT 311
Db 5211 AGTATTGTAGCGTTCTCTATGCTAATAAATTTTGTCTATATATTCAAAATATATAAAA 5152
QY 312 GATGAGTACCACAAATCAAGATGATAGGTGTTTAATCACTGTAAAAATGTTGATTGAATAC 371
Db 5151 TAAAAAGACAAAATTAATGGTTACATTATTAGCTAAATAAATATTTTAACITTTAAAAA 5092
QY 372 TAGCAATGCAGAACATATACATATTTTAACTCTCTTTGGAAATTTTGTGTTTTGTTTAT 431
Db 5091 TTTTATTAAATATAAAAAATTTTAAAAACAATGTTTAAAGAAATATTTTTTTAGAGTTTCT 491
```

Db 5091 TTTTATTTAAATATAAAAAATTTTAAACAATGTTAAGAAATATTTTTTTAGAGTTTCT 5032
QY 432 CATTTTGTGAATACACGAAGAGCTCAGTTATATTTTCATATTTGTATATCAATTTGTTCTTAT 491
Db 5031 TTTGAATCATAGAAGGTTCTTATTTTATGATTAATATTTTGTAAATAGAATTTTTTAAAT 4972
QY 492 TAATC 496
Db 4971 TCTTC 4967

RESULT 15

US-10-303-162-1/c
; Sequence 1, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-162-1

Query Match 9.0%; Score 45; DB 4; Length 11474;
Best Local Similarity 49.0%; Pred. No. 0.56;
Matches 120; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 252 AGTCTACTGTCGAGGCCAAACTGTAATTAACGTTTGAGTTTAGAGAAATATTTTGAT 311
Db 5211 AGTATTGTAGCGTTCTCTATGTCGCTAATTAATATTTGCTATATATTCAAATATATAAA 5152
QY 312 GATGAGTACACAAATCAAGATGATAGGTGTTAATCAGTCGTAATAAATGTTGATGAATAC 371
Db 5151 TAAAAAGAACAAAATTAATGGTTACATATTAGCTAAAAATAAATATTTTAACTTTAAAAA 5092
QY 372 TAGGATGCGAGACATATACATATTTTATCTCTTGGAAATTTTGTGTTTCTTTTAT 431
Db 5091 TTTTATTTAATATAAAAAATTTTAAACAATGTTAAGAAATATTTTTTTAGAGTTTCT 5032
QY 432 CATTTTGTGAATACACGAAGAGCTCAGTTATATTTTCATATTTGTATATGAATTTGTTCTAT 491
Db 5031 TTTGAATCATAGAAGGTTCTTATTTTATGATTAATATTTTGTAAATAGAATTTTTTAAAT 4972
QY 492 TAATC 496
Db 4971 TCTTC 4967

Search completed: July 12, 2005, 10:17:33
Job time : 117.348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 542.443 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929c-1_COPY_1_500
Perfect score: 500
Sequence: 1 aatatttaagttgggttat.....tttgtctatttaattccta 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.6	10.9	50000	21	US-10-706-635-25
2	52.8	10.6	802	19	US-10-767-795-5128
3	51	10.2	13919	18	US-10-240-589C-114
4	51	10.2	158001	17	US-10-211-179-11
5	50.8	10.2	8245	15	US-10-311-455-175
6	50.6	10.1	7072	17	US-10-221-613-348
7	50.4	10.1	5912	15	US-10-311-455-575

c	8	50.2	10.0	8056	20	US-10-473-126-386	Sequence 386, App
	9	50.2	10.0	3673778	16	US-10-312-841-2	Sequence 2, Appl
	10	50	10.0	3673778	16	US-10-312-841-1	Sequence 1, Appl
	11	49.8	10.0	15832	14	US-10-239-676-106	Sequence 106, App
	12	49.8	10.0	15832	15	US-10-311-455-1316	Sequence 1316, App
	13	49.8	10.0	15832	15	US-10-240-453-118	Sequence 118, App
	14	49.8	10.0	34688	19	US-10-433-793-89	Sequence 89, Appl
	15	49.6	9.9	5379	15	US-10-311-455-1649	Sequence 1649, App
	16	49.6	9.9	5379	15	US-10-240-485-129	Sequence 129, App
	17	49.6	9.9	6880	17	US-10-321-613-184	Sequence 184, App
	18	49.6	9.9	8056	20	US-10-473-126-386	Sequence 386, App
	19	49.2	9.8	17131	15	US-10-311-455-1035	Sequence 1025, App
	20	48.6	9.7	5379	15	US-10-311-455-1650	Sequence 1650, App
	21	48.6	9.7	5379	15	US-10-240-485-130	Sequence 130, App
	22	48.6	9.7	5567	15	US-10-311-455-1562	Sequence 1562, App
	23	48.6	9.6	5567	18	US-10-240-589C-82	Sequence 82, Appl
	24	48.2	9.6	5520	15	US-10-311-455-1492	Sequence 1492, App
	25	47.8	9.6	34688	19	US-10-433-793-90	Sequence 90, Appl
	26	47.6	9.5	5464	15	US-10-311-455-1890	Sequence 1890, App
	27	47.6	9.5	5546	15	US-10-311-455-364	Sequence 364, App
	28	47.6	9.5	5546	19	US-10-311-507-92	Sequence 92, Appl
	29	47	9.4	8056	20	US-10-473-126-240	Sequence 240, App
	30	47	9.4	8237	18	US-10-221-714A-528	Sequence 528, App
	31	46.8	9.4	7560	15	US-10-311-455-1195	Sequence 1195, App
	32	46.8	9.4	8056	20	US-10-473-126-240	Sequence 240, App
	33	46.8	9.4	73334	15	US-10-311-455-2098	Sequence 2098, App
	34	46.8	9.4	73334	18	US-10-240-589C-128	Sequence 128, App
	35	46.6	9.3	6794	17	US-10-221-613-133	Sequence 133, App
	36	46.6	9.3	16545	15	US-10-311-455-23	Sequence 23, Appl
	37	46.6	9.3	17674	15	US-10-311-455-1318	Sequence 1318, App
	38	46.4	9.3	58320	19	US-10-322-281-88	Sequence 88, Appl
	39	46.2	9.2	435	19	US-10-767-701-16965	Sequence 16965, A
	40	46.2	9.2	11745	18	US-10-240-453-206	Sequence 206, App
	41	46	9.2	5455	18	US-10-240-589C-54	Sequence 54, Appl
	42	46	9.2	5980	15	US-10-311-455-1162	Sequence 1162, App
	43	46	9.2	6731	15	US-10-311-455-1033	Sequence 1033, App
	44	45.8	9.2	5273	15	US-10-311-455-848	Sequence 848, App
	45	45.8	9.2	17144	18	US-10-221-714A-388	Sequence 388, App

ALIGNMENTS

RESULT 1

US-10-706-635-25
; Sequence 25, Application US/10706635
; Publication No US20050014263A1
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; TITLE OF INVENTION: Li, Yi
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous Cells
; FILE REFERENCE: UP-221CXZ1
; CURRENT APPLICATION NUMBER: US/10706,635
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/662,254
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-25

Query Match 10.9%; Score 54.6; DB 21; Length 50000;
Best Local Similarity 52.9%; Pred. No. 0.81; Indels 0; Gaps 0;
Matches 117; Conservative 0; Mismatches 104;

RESULT 7

US-10-311-455-575

Sequence 575, Application US/10311455

Publication No. US20030143606A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation of Cytosine

TITLE OF INVENTION: Cytosine methylation

FILE REFERENCE: 5013.1014

CURRENT APPLICATION NUMBER: US/10/311.455

CURRENT FILING DATE: 2002-12-16

PRIOR APPLICATION NUMBER: PCT/EP01/07537

PRIOR FILING DATE: 2001-07-02

PRIOR APPLICATION NUMBER: DE 10032529.7

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 2424

SEQ ID NO 575

LENGTH: 5912

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)

US-10-311-455-575

Query Match 10.1%; Score 50.4; DB 15; Length 5912;

Best Local Similarity 52.4%; Pred. No. 2.9;

Matches 111; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 287 TTGAGTTTACAGAAATATTTTGGATGATGAGTACCAACATCAAGATGATAGTGTGTTAAAT 346

Db 2981 TTGGAGAGAATGAAATATTTTCTGTTTTTTTGTGATATTTTGTGTTTGTGTTTAAATGTTTT 3040

Qy 347 CACTGTAAAAATGTTGATTTGAATGATCTACGATGCAGAACATATACATATTTTAAATCTCT 406

Db 3041 TTGTTTTTGTGTGGATATATAAATTTAAATATATTTAAATTAAGGTATTTATTATAT 3100

Qy 407 TTGGAATTTTGTGTTTTGTTTTTATCATTTTGTATACACGAGAGCTCAGTTATATTTTC 466

Db 3101 GTAGTATTTTTTTTTTGTAGATGATTTATTTTGTGAAATATTTAAATAGTTTTATTGTTATAT 3160

Qy 467 ATATTGATATGAATTTGTTCTTATTTAAATCTT 498

Db 3161 TTTTATTATTTTTTTTTTTTTTTTTTTTTTTT 3192

RESULT 8

US-10-473-126-386/c

Sequence 386, Application US/10473126

Publication No. US20040234973A1

GENERAL INFORMATION:

APPLICANT: Epigenomics AG

TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell differentiation

TITLE OF INVENTION: proliferative disorders

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/10/473.126

CURRENT FILING DATE: 2003-09-26

NUMBER OF SEQ ID NOS: 1258

SEQ ID NO 386

LENGTH: 8056

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)

US-10-473-126-386

Query Match 10.0%; Score 50.2; DB 20; Length 8056;

Best Local Similarity 46.0%; Pred. No. 3.5;


```
RESULT 10
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312.841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 10.0%; Score 50; DB 16; Length 3673778;
Best Local Similarity 45.2%; Pred. No. 28;
Matches 226; Conservative 0; Mismatches 270; Indels 4; Gaps 1;

QY 2 ATATTTAAGTTTGGTTTATATCTTCTAGTAATCTTTGAAATATTTGTAAGAGATAATGC 61
Db 1713502 ATATATATATTTTATATATATATATATATATATATATATATATATATATATAT 1713561

QY 62 TTCTAATAAATACATGGATTTTATGGAAATTAATGTAATGTAAGAAACTATGCATAACT 121
Db 1713562 TTTATATATATTTTATATATTTTATATATATATATATATATATATATATATATAT 1713621

QY 122 ACAGTGTATTTTGGACGACCAAAATGATATATATATATATATATATATATATATAT 181
Db 1713622 ATTTTATATATTTTATATATATTTTATATATATATATATATATATATATATAT 1713681

QY 182 AGTGAATAGGATGCGGACA---AGTTGATCGACTCTAAACATTTATGGTACGTAAT 237
Db 1713682 ATATTTTATATATATTTTATATATATATATATATATATATATATATATATATAT 1713741

QY 238 TCCGACAGTGGTTACAGCTACTCTCGAGGCCAACTGGTAATTAACGTTTGAAGTTAG 297
Db 1713742 ATATATATATATATATTTTATATATATTTTATATATATTTTATATATATTTT 1713801

QY 298 AGAATATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
Db 1713802 ATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATAT 1713861

QY 358 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
Db 1713862 TTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATAT 1713921

QY 418 GTTTTGTGTTTATCAATTTTGAATACACGAGAGCTCAGTTATTTTCATATTTGATAT 477
Db 1713922 TATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTT 1713981

QY 478 GAATTTGTTCTATTTAACT 497
Db 1713982 ATATATTTTATATATATTTT 1714001

RESULT 11
US-10-239-676-106
; Sequence 106, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
```

```
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 106
; LENGTH: 15832
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (2778, 2845..2846, 2865, 2916, 9325, 10589, 10612, 10809, 10817)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (10821, 10827, 10830..10831, 10839, 12552, 13433, 15042)
US-10-239-676-106

Query Match 10.0%; Score 49.8; DB 14; Length 15832;
Best Local Similarity 60.9%; Pred. No. 5.2;
Matches 81; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 360 TTGATTGAATACATGCAATGCAACATATACATATATTTTAAATCTCTTTGGAATTTTGT 419
Db 7041 TTTTATATATAGATATATATATATATATATATATATATATATATATATATATAT 7100

QY 420 TTTTGTGTTTATCAATTTTGAATACACGAGAGCTCAGTTATATTTTCATATTTGATATGA 479
Db 7101 TTTTATAGAAATTTTTCGAAAAAATTAATTTAGTTAAATTAATAATTTAGTTTTT 7160

QY 480 ATTTGTTCTATTT 492
Db 7161 TTTATATATATAT 7173

RESULT 12
US-10-311-455-1316
; Sequence 1316, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1316
; LENGTH: 15832
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
```


GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1649
; LENGTH: 5379
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1649

Query Match 9.9%; Score 49.6; DB 15; Length 5379;
Best Local Similarity 47.3%; Pred. No..4;
Matches 212; Conservative 0; Mismatches 234; Indels 2; Gaps 2;

QY 5 TTTAAGTTTGGTTTATATCTCTTAGTAATCTTTGAAATATGTAAGAGATAATGCTTC 64
DB |||||
QY 3553 TATAGGTATGGTTATATATCTCTTAGTAATCTTTGAAATATGTAAGAGATAATGCTTC 64
DB |||||
QY 65 TAATAAATAACATTTGGAATTTATGGAATTTATGTAATTTGAAATACTATGCAAAATACATA 124
DB |||||
QY 3612 TATTATATGCTTACGTTGGTTGGAATTTTGTGATTTTAGTGATTTATTATTATTAGTT 3671
DB |||||
QY 125 GTGTATTTTGGAAAGCAACCAATATATATGTAATCTTCTGTTCTCTCTACATAGT 184
DB |||||
QY 3672 TTTTAAAGTGTGGGATTTATAGGTATGAGTTATTATGTTTGGTTTAGAGATAAGTTTTT 3731
DB |||||
QY 185 GTAATAGATAGCGGACAAAGTTGATCGACTCTAAACATTATGGGTACGTAATTCGCAG 244
DB |||||
QY 3732 TAAATTGATTTTGTATTATTAAGGTGAAGGAAATTTGAAATAATTTTAA 3791
DB |||||
QY 245 TGGTTACAGTCTACTGTCGAGGCCAAACTGTAATTTAAACGTTTGAAGTTTAGAGAAATA 304
DB |||||
QY 3792 TTTTGAATAATAATAAGGATATAAAATTTTAAATTTTGTATTAGAGTTAGT 3851
DB |||||
QY 305 TTTTGATGAT- GAGTACACATCAAGATGATAGGTGTTAATCAGTGTAAATTTGA 363
DB |||||
QY 3852 TATTGTTAATTTGTGGAAGAAATTTTATTTTTGTATATTAATATTATAAATTTTAA 3911
DB |||||
QY 364 TTGAATACGATGACGACATATACATATTTTAACTCTCTTTGGAAATTTTGTGTTT 423
DB |||||
QY 3912 ATAAATTTGATTAATTAATATATATATATAGTTTTCGTTTATTTTATTTTATTTT 3971
DB |||||
QY 424 GTTTTATCATTTTGAATACACGAAGA 451
DB |||||
QY 3972 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3999
DB |||||

Search completed: July 12, 2005, 11:12:08
Job time : 554.443 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:03:44 ; Search time 2463.72 Seconds
(without alignments)
7724.974 Million cell updates/sec

Title: US-10-018-929c-1_COPY_1_500

Perfect score: 500
Sequence: 1 aatatttaagtttggttat.....tttggtctatttaatttca 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	302.4	60.5	421	8 BZ379935	BZ379935 SALK_1143
2	95.6	19.1	146	8 BZ594136	BZ594136 SALK_0829
3	76.4	15.3	1101	9 CNS0039G	AL063921 Drosophil
4	65.6	13.1	1101	9 CNS000EVL	AL063706 Drosophil
5	62.2	12.4	1101	9 CNS003BD	AL064091 Drosophil
6	61.6	12.3	1101	9 CNS000EO7	AL063940 Drosophil
7	59.8	12.0	1101	9 CNS000EVL	AL063706 Drosophil
8	59.8	12.0	1592	9 CG750135	CG750135 P044-3-D0
9	57.6	11.5	1204	9 CNS016E2	AL106628 Drosophil
10	56	11.2	997	6 CD049644	CD049644 AGENCOURT
11	55.8	11.2	973	9 CNS071LE	AL425064 clone BAO
12	55	11.0	928	9 CNS000DKY	AL071865 Drosophil
13	55	11.0	1086	9 CNS000YXK	AL096962 Drosophil
14	54.4	10.9	1539	9 AG340947	AG340947 Mus muscu
15	54.2	10.8	1101	9 CNS000Z6Z	AL097301 Drosophil
16	54	10.8	1101	9 CNS000EO7	AL063940 Drosophil
17	53.2	10.6	1061	9 CNS015LM	AL105604 Drosophil
18	53.2	10.6	1101	9 CNS000Z2U	AL097152 Drosophil
19	53	10.6	945	9 CNS004D0K	AL285149 Tetradon
20	52.8	10.6	524	9 CNS01U90	AL167541 Tetradon
21	52.8	10.6	872	7 CK022958	CK022958 AGENCOURT
22	52.2	10.4	640	8 BHI86548	BHI86548 029_H_16-
23	52.2	10.4	640	9 CNS07R05	AL6223495 T3 end of
24	52.2	10.4	768	9 AG557213	AG557213 Mus muscu

C	25	52	10.4	430	6	CA856066	CA856066 PfESToacs
	26	52	10.4	1164	8	CC218891	CC218891 CH261-14M
	27	52	10.4	1313	7	CK997149	CK997149 ip1sc02.b
	28	51.8	10.4	1254	3	CR720510	CR720510 Tetradon
	29	51.4	10.3	1526	9	AG305152	AG305152 Mus muscu
	30	51.2	10.2	734	9	CNS010MP	AL099163 Drosophil
	31	51.2	10.2	796	1	AV704385	AV704385 AV704385
	32	51.2	10.2	1101	9	CNS000FMC	AL070972 Drosophil
	33	51.2	10.2	1225	9	CNS0161D	AL106171 Drosophil
	34	51.2	10.2	1402	9	AG382069	AG382069 Mus muscu
	35	51	10.2	1101	9	CNS0175J	AL107617 Drosophil
	36	51	10.2	1227	9	AG430010	AG430010 Mus muscu
	37	50.8	10.2	1001	9	CNS0155H	AL105023 Drosophil
	38	50.8	10.2	1392	9	CG757503	CG757503 P052-4-C0
	39	50.6	10.1	1038	9	CNS06L7M	AL403832 T3 end of
	40	50.6	10.1	1101	9	CNS00396	AL063911 Drosophil
	41	50.6	10.1	1101	9	CNS017KE	AL108152 Drosophil
	42	50.4	10.1	886	8	BH177277	BH177277 008_L-22-
	43	50.4	10.1	886	9	CNS07JUX	AL614235 T3 end of
	44	50.4	10.1	913	9	CNS00C20	AL059715 Drosophil
	45	50.2	10.0	734	9	CNS010MP	AL099163 Drosophil

ALIGNMENTS

RESULT 1
BZ379935
LOCUS
DEFINITION
SALK_114344.50.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_114344.50.00.x, genomic survey sequence.

ACCESSION
BZ379935
VERSION
BZ379935.1 GI:25472281
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana

REFERENCE
1 (bases 1 to 421)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the

TITLE
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)

JOURNAL
COMMENT
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

FEATURES
source
1. 421
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_114344.50.00.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

Class: TDNA tagged
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 3' end of AT1G08065.

Query Match	60.5%	Score 302.4	DB 8	Length 421	
Best Local Similarity	99.7%	Pred. No. 6.1e-49			
Matches 303	Conservative	0	Mismatches 1	Indels 0	Gaps 0
QY	197	CGGACAAAGGTTGATCGACTCTAAACATATATGGGTACGTAATTCGCGAGTGGTTACAGTCT	256		
Db	1	CGGACAAAGGTTGATCGACTCTAAACATATATGGGTACGTAATTCGCGAGTGGTTACAGTCT	60		
QY	257	ACTGTCTGAGGCCAAACTGGTAAATTTAAACGTTTGAAGTTTACAGAAATATTTTGATGATGA	316		
Db	61	ACTGTCTGAGGCCAAACTGGTAAATTTAAACGTTTGAAGTTTACAGAAATATTTTGATGATGA	120		
QY	317	GTACCAACAATCAAGATGATAGGTGTTAAATCACCTGTAATAATGTTGATTTGAATACTACGA	376		
Db	121	GTACCAACAATCAAGATGATAGGTGTTAAATCACCTGTAATAATGTTGATTTGAATACTACGA	180		
QY	377	ATCGAGAACATATACATATTTTAAATCTCTTTGGAAATTTTGTGTTTGTATTCATTT	436		
Db	181	ATCGAGAACATATACATATTTTAAATCTCTTTGGAAATTTTGTGTTTGTATTCATTT	240		
QY	437	TTCAATACACGAGAGCTCAGTTATATTTTCATATGTATATGATTTGTTCTATTATTC	496		
Db	241	TTCAATACACGAGAGCTCAGTTATATTTTCATATGTATATGATTTGTTCTATTATTC	300		
QY	497	TTCA 500			
Db	301	TTCA 304			
RESULT 2					
BZ594136					
LOCUS		146 bp	DNA	linear	GSS 07-JAN-2003
DEFINITION		SALK_082960.17.65.x Arabidopsis thaliana	TDNA insertion lines		
		Arabidopsis thaliana genomic clone SALK_082960.17.65.x, genomic			
		survey sequence.			
ACCESSION		BZ594136			
VERSION		BZ594136.1	GI:27534655		
KEYWORDS		GSS.			
SOURCE		Arabidopsis thaliana (thale cress)			
ORGANISM		Arabidopsis thaliana			
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoid II; Brassicales; Brassicaceae; Arabidopsis.			
AUTHORS		1 (bases 1 to 146)			
		Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,			
		Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,			
		Shinn,P., Zimmerman,J. and Ecker,J.R.			
TITLE		A sequence-indexed library of insertion mutations in the			
JOURNAL		Arabidopsis Genome			
COMMENT		Unpublished (2001)			
		Contact: Joseph R. Ecker			
		Salk Institute Genomic Analysis Laboratory (SIGAL)			
		The Salk Institute for Biological Studies			
		10010 N. Torrey Pines Road, La Jolla, CA 92037, USA			
		Tel: 858 453 4100 x1752			
		Fax: 858 558 6379			
		Email: ecker@salk.edu			
		This is single pass sequence recovered from the left border of			
		TDNA. This sequence lies within 300 bases of the 3' end of			
		At1g08065.			
Class:		TDNA tagged.			
FEATURES		Location/Qualifiers			
source		1..146			
		/organism="Arabidopsis thaliana"			
		/mol_type="genomic DNA"			
		/ecotype="Col-0"			
		/db_xref="taxon:3702"			
		/clone="SALK_082960.17.65.x"			
		/clone_lib="Arabidopsis thaliana TDNA insertion lines"			
		/notes="PCR was performed on Arabidopsis thaliana lines			
		each of which contains one or more TDNA insertion			
		elements. The resultant fragment for each line was			

[illegible]

RESULT 4	CNS000EVL	1101 bp	DNA	linear	GSS 04-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC:				
DEFINITION	BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL065706				
VERSION	AL065706.1	GI:4949849			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
	- Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeгава and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain-used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center, can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .				
FEATURES	Location/Qualifiers				
source	1. .1101				
	/organism="Drosophila melanogaster"				

```

/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match      13.1%; Score 65.6; DB 9; Length 1101;
Best Local Similarity 36.1%; Pred.No.0.0067;
Matches 180; Conservative 81; Mismatches 235; Indels 2; Gaps 2;

QY      1 AATATTAGTTGGTGTATATCTTCCTAGTAATCTTTGAAATATTTGAAGAGATAATG 60
       ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB     555 AAWAAATAWATTAAWAATTTAAWAAWWTATATTAATWTATAAATTTWATTAAATAAAAA 614
QY     61 CTTCCTAATAAATAACAATTCGATTTTATTCGAAATTAATGTATCGAAAACCATGCCAATAC 120
DB    615 ATATTTTTTWTATAAANAATTTTTTAATTAATTTTAATTAAT-TATTAATAATWATTWTTWTAAT 673
QY    121 TAGAGTGATATTTGGAAACGACCACCAATATATATGTAAACTTTTCGTTCTAGTCTTCTACA 180
DB    674 TAAATAATTTTAAATATAWAAAAAABAAAAAATWAAWATWATAWATAAATTTAAAAWA 733
QY    181 TAGTGTAAATAGGATCGGACGAAGGTTGATCGACTCTAAACAATATATGGGTACGTAATCC 240
DB    734 ATAAGAATAAATAATWAAWATAATATWATAATATWTTTAAWMAWMTATWAAWATAWATAWA 793
QY    241 CGAGTGGTTACAGTCTACTGTCGAGGCCAACCTGGTAAATTAACAGTTTCAAGTTTAGAGA 300
DB    794 ATAAWAAWAAWATAAATAWATAWATAWATAWAAWAAWATAWATAWATAWAAWAAWAAA 853
QY    301 AATATTTCATGATGAGTACCAACAATCAAAGATGATAGGTGTTAATCACTGTAATAATGCT 360
DB    854 AATTAATATTAATWATAWAAAAAATAWAAWTTTWTTTWTWAAWATAAABAATAWATAW 913
QY    361 TGATTGAATACACTGCAATGCGAACAATACATATATTTTAAATCTCTCTTGGAAATTTTGT 420
DB    914 AAAAATAAAAAAAAAAATAAATAWAAWTTWATATTTTATTAAAAATWTTATWTTWATTWAAW 973
QY    421 TTGTGTTTATCATCTTTTGAATACAGA-AGAGCTCAGTTATATTTCAATTTGTATATCA 479
DB    974 TWTATWTTTATWTTATATATWTTATWATAWATAATTTATTAAWWTATATTTTTTAAWAAWTA 1033
QY    480 ATTGTGTTCTATTTAAATCT 497
DB   1034 ATATATWATAWTTAAWAT 1051

RESULT 5
CONS003BD/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence Tet3 end of BAC #
BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL064091
VERSION
AL064091.1 GI:4941847
KEYWORDS
GSS
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequences ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K08"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 12.4%; Score 62.2; DB 9; Length 1101;
Best Local Similarity 38.3%; Pred. No. 0.031;
Matches 190; Conservative 44; Mismatches 262; Indels 0; Gaps 0;
QY 3 TATTTAAGTTTGGTTTATATCTTCTAGTAATCTTTGAAATATTTGTAAGATAATGCT 62
Db 996 TTTTWTATTAAT 937
QY 63 TCTAATAATAACATGGATTTATTCGAATTAATGATTTGAAAAAACTATGCNAATCTA 122
Db 936 WTAAWTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 877
QY 123 CAGTGTATTTTGGACGACCAAAATGATATATGTAATCTTCTAGTCTCTACATA 182
Db 876 WATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 817
QY 183 GTGTAATAGTAGACGCAAGGTTGATCGACTCTAAACATTTATGGGTAGTAATCCGC 242
Db 816 ATTTAAWAAAWTATWNNATATATAATWATATATTTTATATATAAATAAATWATWT 757
QY 243 AGTGGTTACAGTCTACTGTCGAGGCCAACTGTAATTAACGTTTGAAGTTTAGAGAA 302
Db 756 TTTTATTAATATATTTTAAWAAWATATATATTTTAAWTTTATATATTTTATATTTT 697
QY 303 TATTTTGATGATGAGTACACCAATCAAGATGATAGTGTTATCACTGTAATAAATGTTG 362
Db 696 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 637
QY 363 ATTTGAATCTAGCAATGCGAACAATATACATATTTTATTTTATTTTATTTTATTTT 422
Db 636 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 577
QY 423 TGTATTTATCATTTTGAATACAGGAGCTCAGTATATATTTTATATTTTATATGATG 482
Db 576 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 517
QY 483 TGTCTATTTTATCTTT 498
Db 516 TAAAGGTTTATTTTATTTT 501

RESULT 6
CNS000E07/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
AL069440.1 GI:4949583
VERSION
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> the BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
pi and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 12.3%; Score 61.6; DB 9; Length 1101;
Best Local Similarity 36.8%; Pred. No. 0.04;
Matches 186; Conservative 66; Mismatches 247; Indels 6; Gaps 1;
QY 2 ATATTAAGTTTGGTTTATATCTTCTAGTAATCTTTGAAATATTTGTAAGATAATGC 61
Db 956 ATAATWTTTAT 897
QY 62 TTCTAATAATAACATTCGATTTATTTGGAATTAATGTAATGTAATAAATCTATGCAATACT 121
Db 896 TTTAAWTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 837
QY 122 ACAGTGTATTTTGGACGACCAAAATGATATATGTAATCTTCTAGTCTCTTCTACAT 181
Db 836 WAATTTAT 777
QY 182 AGTGTAAATAGGATAGCGGCAAGGTTGATCGACTCTAAACATTTATGGGTACGTAATCCG 241
Db 776 AAATWTTTAT 717
QY 242 CAGTGGTTACAGTCTACTGTCGAGGCCAACTGGTAATATAAAGCTTTGAAGTTTACAGAA 301
Db 716 AAATWTTTAT 657
QY 302 ATATTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361
Db 656 WTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 597
QY 362 GATTTGAATCTACGA-----ATGCGAACAATATATATATATATATATATATATCTTTGG 415
Db 596 AAATWAAAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 537
QY 416 TTGTTTTTCTGTTTATCATTTTATTTTGAATACACGAAGAGCTCAGTATATTTTCAATTTG 475
Db 536 TTTAAWTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 477
QY 476 ATGAATTTCTCTATTTAATCTTCA 500
Db 476 AWTATATTTTTTTTATATATCTTTTCM 452

Qy	386	ATATACATATTTTAACTCCTTTGGAAATTTTTGTTTGTTTTTATCATTTTTGAATACA	445
Dd	741	TWTTTTATTWTTAATTTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAA	682
Qy	446	CGAAGAGCTCAGTTATTTTCATATTTGTATATGAATTTGTTCTATTTAATCTT	498
Dd	681	ATTATTTAAATTAATWAATTAATTTATAATTAATTAATTAATTAATTAATTAATTAATTT	629
RESULT 8			
CG750135/c			
LOCUS	1592 bp DNA linear GSS 24-OCT-2003		
DEFINITION	P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.		
ACCESSION	CG750135		
VERSION	CG750135.1 GI:37971278		
KEYWORDS	GSS.		
SOURCE	Pristionchus pacificus		
ORGANISM	Pristionchus pacificus		
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus. 1 (bases 1 to 1592)		
AUTHORS	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.		
TITLE	An integrated physical and genetic map of the nematode Pristionchus pacificus		
JOURNAL	Mol. Genet. Genomics 269 (5), 715-722 (2003)		
MEDLINE	22835951		
PUBMED	12884007		
COMMENT	Contact: Sommer RJ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel: 00497071601371 Fax: 00497071601498 Email: ralf.sommer@tuebingen.mpg.de Class: BAC ends.		
FEATURES			
source	I. .1592 /organism="Pristionchus pacificus" /mol_type="genomic DNA" /strain="California" /db_xref="taxon:54126" /clone_lib="Ppa EcoRI BAC Library" /note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."		
ORIGIN			
Query Match	12.0%; Score 59.8; DB 9; Length 1592;		
Best Local Similarity	44.0%; Pred. No. 0.086;		
Matches 219; Conservative 0; Mismatches 278; Indels 1; Gaps 1;			
Qy	2	ATATTTAAAGTGTGGTTTATTTCTTCTAGTAATCTTTGAAAATATTGTAAGAGATAATGC	61
Dd	965	ATAATTAATATAAATAAATAATTTNTATATATATATATATATATATATATATATATATN	906
Qy	62	TTCTAATAAATAACATGGATTTATGGGAATTAATGATGTGAAAAAACCTATGCAATACT	121
Dd	905	TATAAAATATTAATAAATAAATAATTAATAAATNTTATAATAATNAATAATTTATTTAATAA	846
Qy	122	ACAGTGTATTTTGGACGACCACCAATGATATATGTAACCTTTTCGTCTA-GTCTTCTACA	180
Dd	845	AAATTAATAAATTTATTTATNTNTAAANAATTAATTAATAATATATATATATAATAAATAA	786
Qy	181	TAGTGTAAATAGGATAGCGACAAGGTGTGATCGACTCTAAACAAATTTGGTAGCTAATCC	240
Dd	785	AATTATAAATTTATTTANTATATAATTAATTAATTAATTAATTAATTAATAAATATTTA	726
Qy	241	GCAGTGGTTTACAGTCTACTGTGCGAGGCCCAAACCTGGTAATTAATAACGTTTGAAGTTAGAGA	300
Dd	725	ATAATAAATAAATATATAATAATAAAAAAAAAATTTAAAAAATAATTTATAATAAATAATTA	666

```

QY 301 AATATTTTGTATGATGAGTACCAACATCAAGATGATAGGTGTTTAATCAGTCTGTAAATAATGT 360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
665 TAATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 606
QY 361 TGAATGAATACGATGACGAGTACGAGTACATATATTTTAAATCTCTTTGGAATTTTGT 420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
605 AAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 546
QY 421 TTTGTTTTTATCATTTTGAATACACGAGAGCTCAGTTATATTTTCAATATGTTATATGAA 480
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
545 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 486
QY 481 TTTGTTCTATTAATCTT 498
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
485 TNNNTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 468

```

```

RESULT 9
CNS016E2 1204 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106628
VERSION AL106628.1 GI:5622852
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1204)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - France (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

```

```

FEATURES             Location/Qualifiers
     source            1..1204
                        /organism="Drosophila melanogaster"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7227"
                        /clone="BACN15A12"
                        /clone_lib="DrosBAC"
                        /plasmid="pBelobAC11"
                        /notes="end : T7"

```

```

ORIGIN
Query Match          11.5%; Score 57.6; DB 9; Length 1204;
Best Local Similarity 22.5%; Pred. No. 0.24;
Matches 96; Conservative 155; Mismatches 174; Indels 1; Gaps 1;

```

```

QY 1 AATATTAAAGTTGGTTTATATTCTTTCTAGTAATCTTTGAAATATTGTAAGAGATAATG 60
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1203 AWTADAKAAWAATATAKAWKAWKATAAKAWKAKAWKAAKAAKAAKAAKAAKAAKAAKAAKAA 1144
QY 61 CTTCCTAATAATAACATTTGATTTTATTTGGAATTAATGTTATTTGAAACATCTCAATAC 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1143 DAAAAKKGAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 1084
QY 121 TACAGTGTATTTTGGAGACGACCAAAATGATATATGTAATCTTTGTTCTAGTCTTCTACA 180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1083 KWAARKWATKKDDDKDAADAKKKKKAKADADKADKADKADKADKADKADKADKADKADKADK 1024

```

```

QY 181 TAGTGTATAGGATAGCGACCAAGGTGTATCGACTCTAAACATTATGGGTGATGCTAATTC 240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1023 KKKKAKKKKKKKDDADADWAAKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 965
QY 241 GCAGTGGTTTACAGTCTACTGTCTGAGGCCAAACTGGTAAATTAACCGTTTGAAGTTTAGAGA 300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 KKKKKKKKKKKKKKAAKADAAAATAAAKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 905
QY 301 AATATTTTGTATGATGATGATACCAATCAAGATGATAGGTGTTTAATCAGTCTGTAAATAATGT 360
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 DAAADADKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 845
QY 361 TGATTTGAATCTACGAAATCGAACAATATATATATTTTAAATCTCTTTGGAATTTTGT 420
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
844 KAGGAARAATAKGGGGRGGGAGGAAAGATTTTAKTDDTTTKAGADAKTTTTTKT 785
QY 421 TTTGTT 426
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 TTWATK 779

```

```

RESULT 10
CD049644/c
LOCUS CD049644
DEFINITION AGENCOURT_13986201 NIH_MGC_172 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CD049644
VERSION CD049644.1 GI:30485777
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 997)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jamie Thompson, University of WI
cDNA Library Preparation: Gina Zastrow-Hayes
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: NDKM47 row: e column: 22
High quality sequence start: 10
High quality sequence stop: 476.
FEATURES             Location/Qualifiers
     source            1..997
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /tissue_type="embryonic stem cells, WA01, passage 38"
                        /lab_host="DH10B Tona"
                        /clone_lib="NIH_MGC_172"
                        /note="Vector: pDONR201; Site 1: attP2; Site 2: attP1; LIBR PRIMING - oligo dt; METHOD - full-length enriched; Embryonic Stem Cells H1; LIBR PROVIDER - Bradfield"

```

```

ORIGIN
Query Match          11.2%; Score 56; DB 6; Length 997;
Best Local Similarity 45.5%; Pred. No. 0.49;
Matches 225; Conservative 0; Mismatches 268; Indels 2; Gaps 2;

```

```

QY 4 ATTATAGTTGGTTTATATTCTTTCTAGTAATCTTTGAAATATTGTAAGAGATAATGCTT 63
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 ATTAATGATTAATATAAATTTGTTATGTTTAAAAATAAATTTATTTATATATATA 838
QY 64 CTAATAAATAACATTTGGATTTTATGGAATTAATGTTTGAATAAATCTGCAAACTACTAC 123
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
837 AATATAAANAATTTTATTTTATTAATAATAATATATATATAAATAAATAAATAAATAAATAA 778

```


TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammeter in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw ep, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
 1..928
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR27A24"
 /clone_lib="RPCI-98"
 /note="end : T7"

ORIGIN

Query Match 11.0%; Score 55; DB 9; Length 928;
 Best Local Similarity 31.6%; Pred. No. 0.77;
 Matches 138; Conservative 82; Mismatches 217; Indels 0; Gaps 0;
 QY 2 ATATTTAAGTTGGTTTATATCTCTTCTAGTAATCTTGAATAATTGTAAGAGATAATGC 61
 Db 479 WTATATATATAAATAATATATGTAATAATTTTATKAWAAAAAATAATTTTTT 538
 QY 62 TTCTAATAATAACATTTGATTTTATTTGGAATTAATGTAATTTGAAAAAACTATGCAATACT 121
 Db 539 WTTTTTAWAAAAATATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 598
 QY 122 ACAGTGATTTTGGACGACCAAAATGATATATGTAACCTTCTGCTAGCTTCTACAT 181
 Db 599 AWAATTTTATATWAATAAATAATATATATATATATATATATATATATATATATATA 658
 QY 182 AGTGTAATAGGATAGCGGACAGGTTGATCGACTCTAAACATTTATGGTACGTAATCCG 241
 Db 659 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 718
 QY 242 CAGTGGTTACAGTCTACTGTGCGAGGCCAAACTCGTAAATTAACGTTTGAAGTTTAGAGAA 301
 Db 719 TTTTATATWAWATAAATAATATWWWAAATWTWDGKNMNNNAWWWWWWWWAAAA 778
 QY 302 ATATTTTGTAGTAGTACCAATCAAGATGATAGGTGTTAATCACTGTAAAAATGTT 361
 Db 779 WWWWWWWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 838
 QY 362 GATTGAATCACTAGCATGCAACATATACATATTTTATATCTTTTGGAAATTTTGT 421
 Db 839 KKKKGGKKTKTKTTTGAARWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 898
 QY 422 TTGTTTTTATCAATTTT 438
 Db 899 TTTTTTTTTTTTTTTT 915

RESULT 13
CNS00YXK 1086 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN01M03 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
ACCESSION AL096962

VERSION AL096962.1 GI:5608573
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1086)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.

FEATURES Location/Qualifiers
 1..1086
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN01M03"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T7"

ORIGIN

Query Match 11.0%; Score 55; DB 9; Length 1086;
 Best Local Similarity 27.9%; Pred. No. 0.77;
 Matches 139; Conservative 121; Mismatches 238; Indels 1; Gaps 1;
 QY 1 AATATTTAAGTTGGTTTATATCTTCTAGTAATCTTGAATAATTGTAAGAGATAATG 60
 Db 569 RATAWTTTRKTTTWTGTTTWTGKGGGGTTTGAATAATATATATKTKRKGKK 628
 QY 61 CTCTTAATAATAACATTTGGAATTTATGTAATGTAATGAAAAAATATGCAATAATAC 120
 Db 629 ATAAKAAKGGGGGAGAGAGTGTGDAKGAATAGAAAAGGGGAGGGATDAG 688
 QY 121 TACAGTGTATTTTGGAAACGACCAAAATGATATATGTAACCTTTCGTTCTAGTCTTACA 180
 Db 689 RGRKAAAWADKKGDADKFAAAAAAATKGAAGDRAAGATTKATTDGDDGGGGGGRGRGG 748
 QY 181 TAGTGTATAGGATAGCGACGAAGGTTGATCGACTCTAAACATTTATGGGTACGTAATCC 240
 Db 749 GTAAATAGGAAARATTTGGKGGKRAAGKNNNGKKKKKKKKKKKKKKKKKKKK 808
 QY 241 CGAGTGGTTTACAGTCTACTGTGCGAGGCCAAACTGGTAATTAACGTTTGAAGTTTAGAGA 300
 Db 809 GDAARAGADKKKKKKKKKKKKKAAATWKGKAAATTKTNAAKTTTAAAAATTTWAAA 868
 QY 301 AATATTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 Db 869 WWWWTKTWTWTWTWAAAAAATTTTWTATAAATAAAGTTTAAAAAATAATAAATAATAT 928
 QY 361 TGAATGAATCACTACGATGCAACATATACATATTTTAACTCTCTTGAATTTTGT 420
 Db 929 TTTTMTAAAWATWATWATTTTAAWWAAAWATAATTTAAATAATKAWAAATKCKKCKWAT 988
 QY 421 TTGTTTTTATCATTTTGAATA-CACGAAGAGCTCAGTTATATATATTTTATATGATGA 479
 Db 989 AKKWAATWAAAAATAATKWWWWATTDKDTTWTATATAATAATTTKTTATWTTT 1048
 QY 480 ATTGTTCTTATTAATCTT 498
 Db 1049 WTTTTTTTTTWWATKKK 1067

RESULT 14
AG340947
LOCUS
DEFINITION
Mus musculus molossinus DNA, clone:MSMg01-134C16.T7, genomic survey
sequence.
ACCESSION
AG340947
VERSION
AG340947.1 GI:47914257
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mus musculus molossinus
REFERENCE
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 1539)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Teukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : DBACE3.6
R.Site 1 : ECORI.
R.Site 2 : ECORI.
COMMENTS

FEATURES
source
1. 1539
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-134C16.T7"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 10.9%; Score 54.4; DB 9; Length 1539;
Best Local Similarity 46.9%; Pred. No. 0.99;
Matches 235; Conservative 0; Mismatches 262; Indels 4; Gaps 2;
QY 4 ATTTAGTTGGTTTATATCTTCTAGTAATCTTTGAAATATGTGAAGATAATGCTT 63
DB 172 ATTTAATTTTCTTTAAATTTATTTTATTTTATTTTATTTTATTTTATTTAAT 231
QY 64 CTAATAAATCAATTCGATTTA---TTGGAATTAATGTATTGAAACAACTATGCAATAC 120
DB 232 TTAATTAATTAATTTTATTAATTAATTAATTAATTTTATTAATTAATTAATTTTAT 291
QY 121 TACAGTGTATTTTGGAAACGACCAAAATGATATATGTAACCTTTGGTTCTAGTCTTCTACA 180
DB 292 TATAATTTAATTTAATATAAATTTTAAATATTTTATATATTTATTTATTTTATTTAT 351
QY 181 TAGTGAATAGGATAGCGGCAAGTTGATCGACT-CTAAACATTAATGGGTAGCTAATTC 239
DB 352 TATTTTATTTAATAATATATATATATTTTATTTTATTTAATTAATATATATTTTATTA 411
QY 240 CGAGTGGTTACAGTCTACTGTGAGGCCCAACTGGTAATTAACGTTTGAAGTTTAGAG 299
DB 412 TATAAATTTTTTTTTTTTAAATANAATTAATTAATTCATTATAATATTTTATTTATTA 471

QY 300 AATATATTTTCATGATGAGTACCACATCAAGATGATAGGTGTTAATCACTGTAATAAATG 359
DB 472 TATTTTATTAATCAATTTAATAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 531
QY 360 TTGATTGAATACACTACGAATGCAGAACATATACATATTTTAAATCTCTTTGGAATTTTGT 419
DB 532 TTATTTATATATAATTTTAAATTTTATTTAATTTTATTTTATTTTATTTAATTTATTT 591
QY 420 TTTTGTGTTTATCATTTTGAATACAGAGACTCAGTTTATATTTTTCATATTTGATATGA 479
DB 592 ATTTTATTTTATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 651
QY 480 ATTTGTTCTATTTAACTTCA 500
DB 652 ATTTATTTTATTTTAAATTTTA 672

RESULT 15
CNS0026Z/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN01A10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
AL097301 GI:5608912
VERSION
AL097301.1
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
TITLE
Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN01A10"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : T7"

ORIGIN
Query Match 10.8%; Score 54.2; DB 9; Length 1101;
Best Local Similarity 26.6%; Pred. No. 1.1;
Matches 130; Conservative 131; Mismatches 224; Indels 4; Gaps 1;
QY 10 GTTTGGTTTATATCTTCTAGTAATCTTTGAAATATTTGTAAGAGATAATGCTTCTAATA 69
DB 1097 GTWATTTGWRWMTDADMKTTDTTWTWTTWRTADWRKRTGGWRARATKKWWTAKAGAR 1038
QY 70 AATAACATTGGATTTATTCGAATTAATGTATTGAAACAACTATGCAAACTACTACAGTGA 129
DB 1037 AWARTGAGAGAGAGAWGTGDKDDGVKRGGAARAAAKAKWGTAWADGKTWTAHTK 978
QY 130 TTTTGGAAACCAACAAATGATATATGATTAATCTTTCGTTCTAGTCTTCTACATAGTGAAT 189
DB 977 KTWAGWAWAWATTAANKWAKVRWMMTTTWTWMDHDTGTTTTHETKDKKTRNM 918
QY 190 AGGATAGCGCAAGGTTTGTATCGACTCTAAACATTTATGGGTACGTAAATCCGAGTGGTT 249

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2534.52 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929c-1_COPY_4900_5400
Perfect score: 501
Sequence: 1 aatagtttaattctgctcc.....tggcttcgtaccatgcac 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_atg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10329	6	AX063582
2	501	100.0	10329	8	AF213628 Arabidops
3	501	100.0	120965	8	AC026875 Genomic s
4	275	54.9	695	8	AT552638 Arabidops
5	244.6	48.8	6584	8	AF213627 Arabidops
6	244.6	48.8	6571	6	AX063583
7	60	12.0	467	6	AX063609 Sequence
8	43.8	8.7	73009	8	AC007069 Arabidops
9	43.8	8.7	259375	2	AC102623 Mus muscu
10	42.6	8.5	165075	5	BX248091 Zebrafish
11	42.6	8.5	177435	2	CR361544 Danio rer
12	41.8	8.3	146887	9	AL591624 Human DNA
13	41.8	8.3	178384	9	AL445250 Human DNA
14	41.8	8.3	178951	2	AC067897 Homo sapi
15	41.6	8.3	139118	9	AL365315 Human DNA
16	41.6	8.3	169972	2	AC018827 Homo sapi
17	41.6	8.3	138101	2	AC066596 Homo sapi
18	41.4	8.3	13629	3	AF278826 Plasmodiu
19	41.2	8.2	172151	9	HSAC000114 Human PAC

20	41.2	8.2	225603	9	CNS01RG6
21	41	8.2	176837	2	CR392355 Danio rer
22	40.6	8.1	129120	9	HS187B23 Human DNA
23	40.4	8.1	41001	2	AC144490 Homo sapi
24	40.4	8.1	209378	2	AC151545 Callithri
25	40.2	8.0	3357	6	CQ652756 Sequence
26	40.2	8.0	3537	6	CQ644204 Sequence
27	40.2	8.0	3540	6	AX607208 Sequence
28	40.2	8.0	18415	6	AX602174 Sequence
29	40.2	8.0	21253	1	AE014224 Streptoco
30	40.2	8.0	159066	9	AC103557 Homo sapi
31	40.2	8.0	224050	1	SAG766846 Streptoco
32	40.2	8.0	349980	6	CQ655069 Sequence
33	40.2	8.0	349980	6	AX954529 Sequence
34	40	8.0	785	9	HS329087 Homo sapi
35	40	8.0	2099	9	BC039106 Homo sapi
36	40	8.0	164394	9	AC008706 Homo sapi
37	40	8.0	186451	10	AC122211 Mus muscu
38	40	8.0	189846	10	AC147234 Mus muscu
39	40	8.0	237968	2	AC113776 Rattus no
40	39.8	7.9	1766	5	AF510105 Oreochrom
41	39.8	7.9	1771	5	AF510104 Oreochrom
42	39.8	7.9	150197	2	CR394568 Danio rer
43	39.8	7.9	196341	2	BX293552 Danio rer
44	39.8	7.9	248002	5	BX323588 Zebrafish
45	39.6	7.9	120815	10	AL831714 Mouse DNA

ALIGNMENTS

RESULT 1
AX063582
LOCUS AX063582 10329 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100801.
ACCESSION AX063582
VERSION AX063582.1 GI:12541306
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE 1
AUTHORS Habu,Y., Mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 1 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft m.b.H. (AT)

FEATURES
source Location/Qualifiers
1..10329
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

Handwritten signature

Query Match 100.0%; Score 501; DB 6; Length 10329;
Best Local Similarity 100.0%; Pred. No. 1.8e-117;

Matches	501; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	1	AATAGTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTTATGCAACTACTTTGCTG	60		
Db	4900	AATAGTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTTATGCAACTACTTTGCTG	4959		
QY	61	ACTATCTTTCTTACAGGGGCTTTGGAGTCCACAAAAGTCATAAACACTCCAGTCATCATGT	120		
Db	4960	ACTATCTTTCTTACAGGGGCTTTGGAGTCCACAAAAGTCATAAACACTCCAGTCATCATGT	5019		
QY	121	AAATCTTCTGGTACAGATGGTACATGGATGGAATGATGCTTTTGGCTTGTATTCTATG	180		
Db	5020	AAATCTTCTGGTACAGATGGTACATGGATGGAATGATGCTTTTGGCTTGTATTCTATG	5079		
QY	181	GGCAGCCATATCTCTGGAATCCACAGAGGATATGTTAGCTAGTCAAGATTCGGGAAAAATA	240		
Db	5080	GGCAGCCATATCTCTGGAATCCACAGAGGATATGTTAGCTAGTCAAGATTCGGGAAAAATA	5139		
QY	241	CCGATGAATCAACAGAGGAGGCTCCACACTGTTTTAAAGCCGAAAGATGCAAAACTTTGTC	300		
Db	5140	CCGATGAATCAACAGAGGAGGCTCCACACTGTTTTAAAGCCGAAAGATGCAAAACTTTGTC	5199		
QY	301	CAAGTTTTCGATCTTTCAGTAAGTGCCCTTTTTCACCTCCACAACTTATTATGCTTGC	360		
Db	5200	CAAGTTTTCGATCTTTCAGTAAGTGCCCTTTTTCACCTCCACAACTTATTATGCTTGC	5259		
QY	361	ATATGCTTATATATACCTGATGCAACTGTAGTGTGTTACTGATTTCTCTGTTACAGCAA	420		
Db	5260	ATATGCTTATATATACCTGATGCAACTGTAGTGTGTTACTGATTTCTCTGTTACAGCAA	5319		
QY	421	ATGTGAGAGTTTATTCTTCAACTATATATCCATCCGTTTAAAGCATATTTTATTCTTATAT	480		
Db	5320	ATGTGAGAGTTTATTCTTCAACTATATATCCATCCGTTTAAAGCATATTTTATTCTTATAT	5379		
QY	481	CTGGCTTCGTTACCAATGCAC	501		
Db	5380	CTGGCTTCGTTACCAATGCAC	5400		
RESULT 2	AF213628	10329 bp	DNA	linear	PLN 01-JUN-2000
LOCUS	Arabidopsis thaliana MOM gene, complete cds.				
DEFINITION	AF213628				
ACCESSION	AF213628.1	GI:8132769			
VERSION					
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucotsids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 10329)				
AUTHORS	Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.				
TITLE	Disruption of the plant gene MOM releases transcriptional silencing				
JOURNAL	Nature 405 (6783), 203-206 (2000)				
MEDLINE	20279299				
PUBMED	10821279				
REFERENCE	2 (bases 1 to 10329)				
AUTHORS	Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland				
FEATURES	Location/Qualifiers				
source	1. 10329				
	/organism="Arabidopsis thaliana"				
	/mol type="genomic DNA"				
	/cultivar="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/map="between ATEAT1 and phyA"				
mRNA	join(705..1008,1296..2550),2674..2752,2868..3113,				


```
gb|A1012185.1, emb|Z17990.1, gb|AW267999.1, gb|N96009.1,
dbj|AV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"
/translacion="MGKEXFHINIVVIGHVDSKSTTTHLIYKLGIDKRVIERPEK
EAAEMNRKSFVAMVLDLKAERGITIDIALWKPEFTKYCTVIDAPKRDVFIKNM
ITGTSQADCAVLIIDSTGGFEAGISKDGOTREHALLATLGVKOMICCNKMDATTP
KYSKARYDIIKEVSSVLKGVNPDKIPEVPISGFEGDNMLERSNLDWYKGPITLLE
ALDQINEPRSPDKPRLPLQDYKIGGIGTVPGRVETGMKPKPMVTFAPGLTTE
VXSEVHSHSLLEALPDGNDVGNVAVKDLKRGVVASNSKDDPANGAANFTSVYII
MHPGGIGNGYAPLVDSHTIAVFESEILTKIDRSRGEIEKEPKPLKNGDAGVMYK
TPTKPMVTFSGYPLGRFAVDMRQTVAVGVKISVDKDPGTAKVTKAAVKKGAK"
complement (20787..20857)
/product="tRNA-Lys"
/notes="codon recognized: AAA; putative tRNA-Lys"
/evidence=not experimental
complement (Join(21956..22189,22535..22600,22713..22929,
23041..23126))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translacion="MCLNRLPSATPIPRAEERECECMGFNRHOLLSSFLCKRPA
IIMNKGSGSGSGSGSPTAAAAAALQKALLORVETDITSVDNFTQVNVSRV
SPPVKNQSTTNMWRASRMOQAOLSLKLVSSELQATIFSGFASINDHVEQRIEEF
DQEAETNLLARIADDAANLKELESYSSAQRLLDI"
join (23576..23689,24080..24258,24347..24522,24635..24726)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translacion="MTLGLARLVAPMILLFPIELVKAIVITLTPETFSKIALSN
KQALONGCISWCSFYSLSDDTTCLLSPCLLLVQIEKDTAMVFCVPCWCKCKKLG
NWEDLGKAMEGDEIEVGEVDCGTBRVCTKVEIHSYPTFMLFYNGEVSRYKGRD
VSLKAFVVEETEKAQKQLEDEKEL"
complement (Join(25423..25563,25675..26048,26127..26330,
26585..26783,26941..27842,27936..27991,28208..28413))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translacion="MDSVRRKINNTTQSSPSPKSFVYRNPALAAASTANSIRP
SKSVLPIFLISFASAFSLISFMAGEKWLNTALTFKISQEAAYTVTKAOGVLVLCF
IGMVALSGISHRAKPAARGTKSFSEKTDQFSLNRQLELLGIKKQDSUSEF
PKSRPAKPAQSLIEPLVPIHHQTLTSAHSGSGGKLNRSQSLSSSTSPSKQMS
PSMYLYPSSVSSNLISQDQKAVSPWMSGRSSAKDITTEQLQSLARIDEKITE
SAGKMTPTPTVSGFAMASPTVGTSGASGATRPLPVRMSPGAQFTPTPPKQES
GDPPNMSLERAIEGFIHLGVPOIEDWRDLRQWCSVSLPLKLNKVTSHIQVMQT
ASKLVNVTVSQVSDLPNTGATATLAVDRTKGWPYSYSLDEDLALLHOLRANLQAI
DSMQKLTENQGSQQQQQQQQAALIIPWQECVDAISEHRLQSLMGKGEWVKLLP
RSIPADYTVQIRALAEGTCKVNIYNGADTRKNKWSLEPFDSDSHLLLYLFCF
LEHPKWLHLDPSYTGTKASKNPLFLGVLPPKRPPEKVIYAVSGVPSPTLHPGCVL
AVDQSPPTFALYWDKKVQPTLQGRLTALWDSMLLICHRIVKGYGVVGRGNLGSALN
ILQVDSDDDD"
complement (Join(29418..29491,29579..29620,29858..29942,
30386..30805))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translacion="MVSSKRPKEKKARSDDVVVNKASGRSKRSRSGSTTKTSKNVIVK
KPETIYEIIESSSSDSVEAIRGDEAKNGVGVVSKRNGKSVGIPTKSKNREDDGG
```

```
ABDAKIKFPMRIRIRIMRSDNSAPOIMQDAVFLVKNKATEMPIERSEAYDSSVKDKK
KPIHYKHLSSVVNSDQRYEFLADSVPEKLKAEALDEEWERGMTDAG"
Query Match 100.0%; Score 501; DB 8; Length 120965;
Best Local Similarity 100.0%; Pred. No. 1.3e-117;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATATGCAACTACTTTTGGCTG 60
DB 62784 AATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATATGCAACTACTTTTGGCTG 62725
QY 61 ACTATCTTTCTACAGGGGCTTTGGAGTACCAAAAAGTCATACCACTCCAGTCATCATGT 120
DB 62724 ACTATCTTTCTACAGGGGCTTTGGAGTACCAAAAAGTCATACCACTCCAGTCATCATGT 62665
QY 121 AAATCTTCTGGTACAGATGATACATTTGGATGGAATGATGCTTTTGGCTTTGTTATTCTATG 180
DB 62664 AAATCTTCTGGTACAGATGATACATTTGGATGGAATGATGCTTTTGGCTTTGTTATTCTATG 62605
QY 181 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGTCTAGTCAAGATTTGGGGGAAAAATA 240
DB 62604 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGTCTAGTCAAGATTTGGGGGAAAAATA 62545
QY 241 CCGGATGAATCAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCGCAAACTTTTGC 300
DB 62544 CCGGATGAATCAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCGCAAACTTTTGC 62485
QY 301 CAAATTTTGCATCTTTCACTAGTAGTGGCCCTTTTCACTCCACAACTTATTTAGCCTTGC 360
DB 62484 CAAATTTTGCATCTTTCACTAGTAGTGGCCCTTTTCACTCCACAACTTATTTAGCCTTGC 62425
QY 361 ATATGCTTATATATAGTCACTGAACTGATGCTAGTGTGTTACCTGATTTCTGTTACAGCCAA 420
DB 62424 ATATGCTTATATATAGTCACTGAACTGATGCTAGTGTGTTACCTGATTTCTGTTACAGCCAA 62365
QY 421 ATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 480
DB 62364 ATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 62305
QY 481 CTGGCTTCGTTACCAATGCAC 501
DB 62304 CTGGCTTCGTTACCAATGCAC 62284
RESULT 4
ATH552638 695 bp DNA linear PLN 29-MAR-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, left border, clone
DEFINITION 340812.
ACCESSION AJ552638
VERSION AJ552638.1 GI:29368785
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chauvin, S., Bechtold, N., Cruaud, C., Derose, R., Pelletier, G.,
Lepiniec, L., Caboche, M. and Lecharny, A.
T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)
JOURNAL 22363535
MEDLINE 12446565
PUBMED 2 (bases 1 to 695)
REFERENCE 2
AUTHORS Balzerque, S.
Direct Submission
TITLE Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
JOURNAL PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
```

the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplatane' (<http://www.genoplatane.com> and <http://genoplatane-info.infobiogen.fr>).

FEATURES

source
1. .695
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="wassiliewskijia"
/db_xref="taxon:3702"
/clone="340E12"
/notes="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
1. .695
/notes="T-DNA flanking sequence
left border"

ORIGIN

Query Match 54.9%; Score 275; DB 8; Length 695;
Best Local Similarity 100.0%; Pred. No. 1.4e-59;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATAGTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTGATGCAACTACTTTGCTG 60
Db 421 AATAGTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTGATGCAACTACTTTGCTG 480
QY 61 ACTATCTTCTACAGGGCTTTGGAGTACCAAAAGTCATACACATCCAGTCATCATGT 120
Db 481 ACTATCTTCTACAGGGCTTTGGAGTACCAAAAGTCATACACATCCAGTCATCATGT 540
QY 121 AATCTTCTGTCACAGATGTTACATTCGATGGAATGATGCTTTGGCTTGTATTCTATG 180
Db 541 AATCTTCTGTCACAGATGTTACATTCGATGGAATGATGCTTTGGCTTGTATTCTATG 600
QY 181 GCGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 240
Db 601 GCGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 660
QY 241 CCGATGAATCAGAGAGGCTCCACACTGTTTT 275
Db 661 CCGATGAATCAGAGAGGCTCCACACTGTTTT 695

RESULT 5

AF213627
LOCUS
DEFINITION
Arabidopsis thaliana MOM mRNA, complete cds.
ACCESSION
AF213627
VERSION
AF213627.1
KEYWORDS
GI:8132767
SOURCE
Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustoids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 6554)
Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.
Disruption of the plant gene MOM releases transcriptional silencing
of methylated genes

Nature 405 (6783), 203-206 (2000)

JOURNAL

20279299

MEDLINE

10821279

PUBMED

2 (bases 1 to 6554)
Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and

Paszkowski, J.

Direct Submission

Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box

2543, Basel 4002, Switzerland

Location/Qualifiers

FEATURES

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

source

1. .6554
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Zhuerich"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATGAT1 and phyA"
310. .6315
/function="involved in the maintenance of transcriptional
gene silencing"
/codon_start=1
/product="MOM"
/protein_id="AAF73380.1"
/translation="MKDEKIGLTGRTIYTRSLAASIPASVEQETPLGLRRSSRGTPST
KIRKGTSPDIEORSDSVESTDKIPIMARSYRALFRGLKESALKEVSLQSSKGSN
VTPASATRKSRSLAPASVSKGGIKVNSRTPSSLRNRGKTKEVSLQSSKGSN
VVGCSRRIPAGNDVQGGTDCPPADAGSKRLPVDETSIDKGTDPPLKSVTEKIV
LDASPIVETGDSVIGSPSENLETQKLDGKTDCCSPANAEKTLPVGETSLEKYPQ
KFQDNTDCLPPANAEKRLPVGETSLEKTDLPKSTTETGKVLVASTVETRRDS
VICSPSTNLETQKLVSKTGLETDIVLPLKRRDRTABEILDACATVANGDDHVNSSDG
VIPSQCKNDNRPEMCTCKRQKVGDCNRSVCSIVQPVVEESDNVQDMKETGP
VTSREVEENGQIOHGKSDPKFYSSVPEVMVPVQSLDVOLEYCOTLFSKLSLSL
SKIDLCALBETLNSVRKTDHPYMDASLKQLLTKLELHEILDVEIKASGKLHLHK
MLTHIKKGLKAVFYQATQTPPEGLLNLIEDFVGQRFKPSYHEGIYSSKKNSAIN
NFNKSQCCVLLLETRACQTIKLRADAFILFGLSNLPSHDVHKIKIESCSERT
KIPRLYSVCTVEKALILARQNRQKAVENLRSLSLTHALLMGASVLPKLDHFSHS
ETPDSGVSEFQIMDGIHEFSSILSGKEENEVKLLLEAKHAQGTYSSTLFG
EDHILKSDRESPIFWSKLLGGKPNMKVPSDTQPNRKRVOYFEGSEASPKTGDGN
AKRKASDDVDTPRVDPPVDDDERKASGKHGALSKSVKITLQSSCKSGTDTGL
DGNDAFLGSMGSHISGIPEDMLASQDMGKIPDESQRRLHLVLRPMKAKCQVHLSD
ACTVMGNFLEVIENRIYEEAPATFQAIASLWIAALLKQILSHKESLVANSE
LAFKCRVEVDYIYLSLCKMSLFLEHTQQLQFCFTNSKQSVKTKVQVQKLVQSH
VAFKINTKTSMRNSDECMTEKRSKSHYSTATREDIKTISGIKKVKQVQKLVQSH
BEKMKLLNADKOKLETSSVEAAVIRITCSRTSTOVGDILKLDHNYERKDETK
SEKNECLSLQOMHEVAKKLADEACINIRIKSAAKLKVCVPIOSGNSVSGSSN
ISQNAFQICNNAVEATYADTNCAKSNQVPEAENLTGTMGSGSTQVHEMVDVR
NOETMDVSALRQLTKSOSNEHASITVPEILIPADQCBEFAALNHLSDNCDRT
SAASDDVSRSVPEVSQSLNENSAFSLNREALVTTNRRTSHTVGTDTNLDQ
NREDCSLDQBIEPDAMPVQHLASVVEVTRGAESDOYDODICPMPSLACKQPPAAN
TSENLLEAIEPOSAGSETVETDFAASHQGDVCTPILLSPTGNOPAPANTEGONI
NTSAEPHAGPADVESGDYAVIQETMGQAQDACSLSGSGVGTQSDLGANEGQNTTV
AQLTDGSDAVVTGSDQCADSPMLPSPPGNHPTAVNIEGUNTQSVAEPLHS
GSDACMEISEFQVQVETFANLPHFEGVHSGVATVPSLNNGTGIAVQVPPQ
IPFPVNDPFLHELEKLRESESKTFFBKSLKAELEKMAEYQAEPRRPFHEVE
AEHNTRTTKIEKDKNLVIMNKLANAPLSKTDKVSPPGAPRGKIQOLAAQVSA
LRYNIAPQOLQASSPPALVSPALQOQSSPPAPGAPLQPOASPPSPSVPSALL
LNFVCPMPQRPQPLISNIATPSPVTPATNPLRSPAPHLNSYRPSSTPVATATPTS
SVFPQALTYSAVSIOQQOQPOQSSUSSQSNNEVVVCLSDDE"

ORIGIN

Query Match 48.8%; Score 244.6; DB 8; Length 6554;
Best Local Similarity 98.4%; Pred. No. 6.1e-52;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 69 TCCTACAGGGGCTTTGGAGTCACAAAAGTCATAACATCCAGTCATCATGTAATCTTC 128
Db 2973 TCACATGGGGCTTTGGAGTCACAAAAGTCATAACATCCAGTCATCATGTAATCTTC 3032
QY 129 TGTACAGATGTTACATTTGGAAATGATGCTTTTGGCTTGTATCTATGGCAGCCA 188
Db 3033 TGTACAGATGTTACATTTGGAAATGATGCTTTTGGCTTGTATCTATGGCAGCCA 3092
QY 189 TATCTCTGGAATCCAGAGATATGTTAGCTAGTCAAGATTGGGGGAAATATCCGGATGA 248
Db 3093 TATCTCTGGAATCCAGAGATATGTTAGCTAGTCAAGATTGGGGGAAATATCCGGATGA 3152
QY 249 ATCAGAGAGAGGCTCCACACTGTTTTAAAGCCGAGATGGCAAAACTTTTGCACAGTTTT 308
Db 3153 ATCAGAGAGAGGCTCCACACTGTTTTAAAGCCGAGATGGCAAAACTTTTGCACAGTTTT 3212
QY 309 GCATCTTTCAG 319
|||||

Db 3213 GCATCTTTCAG 3223

RESULT 6

AX063583

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1

Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.

Gene involved in epigenetic gene silencing

Patent: WO 0100801-A 2 04-JAN-2001;

Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft m.b.H. (AT)

Location/Qualifiers

1..6571

/organism="Arabidopsis thaliana"

/mol_type="unassigned DNA"

/db_xref="taxon:3702"

310..6315

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC25098.1"

/db_xref="GI:12541308"

/translation="MKDKDKIGLTGRTVTRSLAASIPASVEQETPGLRSSRGTPT KVTPTASATRKERLASPVSKSGSVKNSPTSSSRNQRKGTVEVLSQSGSDN SIRKGTSPDIDQRKDSVEESNDKIKPIMRSARYALFRGKLKESEALVDASNEEEL VVGCRRIRIPAGNDVQGGKTCPPADAGSKPLPVEDTSLDKGIDFPKLSVETEKIV LDASPIETGDSVDSSEENETKQJQDKGDCSPANAESKTLPPVETSLEKEYIQ KFDQNTDCLPPANAESKLPVGETSLEKIDTDFPLKSTETGKMVLVASPIVETRDD SFCPSPTNLETOKLIVSKTGLETDIVLPLKRRDTAEIELDACATVANGDDHVMSSDG VTPSPGCKNDNRPEMNCNCKKRVKNGDQNRVSCIVOPVEESDNVDIMKETGP VTSREYENGQIQHKGKSPDPKPYSSVVPYAVPVQLSDVQLQYCOITLPSKLSLSL SKIDGALSETLNSVRKTDCHPYVMDASLKQLLTNLEHILIDVKEIKASGLHLDK MLTHIKKGLKAVFYQATPTGGLLGNLEDFVQGPSPKSPYEHGIIYSKGNALN NFNKESQCCVLELLETRACQTIKLLRADAFILFGSSLPNSHDVHKVKEIKESCERT KTFPLSVQTEBEKALILARQNRQKAVENLRSLTHALLMWSGASYLFDKLDHPSHS ETPDVSQVSEQITMDGVIHEFSSILSSKGEENEVKLLLEAKHAQCTSSSDSTLFG EDHI KLSDESPNI FWSKLLGKGNPMWKYPSDTPORNKRQVYEPGEASPKGTGDGN AKRKAADVDVTPDTPVDDDERKAKSGKDHGALSPKAVITLQSKSGSDGTGL DGNDAFLGSMASHGISGIFEDMLASQDWGKIPDESQRRLHVLKPKAKLCQVHLSD ACTSMVGNFVYIENHRIEESPATTFQAFQALSWIAALILVQLSHKESLVFRANSE LAFPKSRFEDVYIYILSCMKSLFLEHTQGLQFCFGNKSQSVWSTKLNVNESLUGAT VRDEKNTKMSNSDEDECMTEKRCSHYSTADRIEIKTISGIIKKYKQVQKLVQEH EEKMKELLNMYADKKQKLSKSVAAVIRITCSRTSQVGDLLKLDHNYERKFEDEIK SEKNECLKSLQMHVEVAKKLAEDACENIRIKQVAAKLCVCPITQSGNNKHFGSSN ISQNA PDVQICNNANVEATYADTNCMAKVNQVPEANTLCTMSGSGSTQQVHEWVDVR NDTMDVSALRSERQLTQSGNSHASITPEIILIPADQCEBPAALNVHLSEQNCDRIT SAASDEDSVSRVPEVSQLENIASPEFSLNKEALVTTENRSTSHVGFQTDNLDDQ NREDCSLDQEIPELAMPVQHILASVYETRGAEISQYQDQICPMPSLQAGKQDPAN TSENLEEEAIEPQASGETVETDFAAQHQGDQVTCPLISSPTGNQPAPEANIRGQNTV NTAEPHVAAGPADVSGDYAVIDQETMGAQDACLPSPSGVGTQSDLGANTBQGVTV AOLPTDGSNAVTPGSPVSDQADASPMPLSSPCGNHPDITAVNIEGLDNTVAEPHIS GSDACEMEISEFGPQVQERSTFANLPHGGVEHSAGVTALVPSLLNNGCTEOLAVOPVQ ITPFVNDPFLRLEKLRSENSKKTIFEEKKSIILKASLERKMAEVQAEFRKKFHEVE AENRTTIKEDKNLNIWNKULANAFUSKCTDKKSPSGAPRGRIQQLGAQVSA LERNYIAPOQLQASFPAPALVAPLQOQSFPAPGPAPLQPOASFPSSVRSPSALL LNPVACPMFPQRPQLISNITAPTSVPTPATNPGLRSPAPHLNSVPSSTPTVATPTS SVPPQALTYSAVSIQQOQEQPQSQSLSSGLQSNNEVWCLSDDE"

48.8%; Score 244.6; DB 6; Length 6571;

Best Local Similarity 98.4%; Pred. No. 6.1e-52;

Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

69 TCCTACAGGGCGCTTTGGAGTCCACAAAGTCATAAACACTCCAGTCATCATGTAAATCTTC 128

AUTHORS Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Shen, M., VanNken, S.E., Barnstead, M.E., Mason, T.M., Bowman, C.L., Ronning, C.M., Benito, M.-I., Carrera, A.J., Creasy, T.H., Buell, C.R., Town, C.D., Nierman, W.C., Fraser, C.M. and Venter, J.C.
JOURNAL REFERENCE Unpublished
AUTHORS 2 (bases 1 to 73009)
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
AUTHORS 3 (bases 1 to 73009)
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598658.
 On Aug 22, 1997 this sequence version replaced gi:1808697.
FEATURES The sequence of BAC R19P19 from Arabidopsis thaliana chromosome 1.
 Location/Qualifiers
 source
 1. .73009
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="m1320"
 /clones="T23K3"
 complement(1. .5590)
 /note="overlap with BAC clone T8011 (AC006069:1. .5590)."
 1
 /note="9434 bases before this point were not included in the submitted sequence, due to overlap with another BAC complement(743. .791)
 /rpt_family="AT_rich"
 1611..1655
 /rpt_family="AT_rich"
 complement(2382..2404)
 /rpt_family="AT_rich"
 2905..2944
 /rpt_family="AT_rich"
 complement(3809..6826)
 /gene="AT2g01820"
 /note="synonym: T23K3.1"
 complement(join(<3809. .4353,4540. .>6826))
 /gene="AT2g01820"
 complement(join(3809. .4353,4540. .6826))
 /gene="AT2g01820"
 /codon_start=1
 /product="putative receptor-like protein kinase"
 /protein_id="AAD21776.1"
 /db_xref="GI:4522003"
 /translation="MNSHLGTLCTFIISLLGLANFSLSQTLDDSTWQSLKSLNLTSDVMSNPCKWQVCGDSNRVTIKIOLKOKGIRGTLPNLQSLSELVILEFLNRLSGPDLGLSELQTLNLDLFTSVKPLFGSMSSLOEMYLENNPDPWVPTVKEA TSQNLTLSSIGIKTDFPGQSLPSLNLKSLQNGLELFPMSFAGTYSISLFLN QKINGSILVGNISLTVLSVLSQNGFSGIPDLGLSLRVFNRENQLTGVPSQL VSLSSTLVNLTNNLYLQGPLKPLKSGVGVIVNNRKFCTNVAGEACDPRVDLTLSVA ESFGYVKLAESWKNPCVNVWIGITCSGGNITVNNRKDLSCTISPSLAKLTSLPT INLADNKLSGHIDELTSLKRLLDVSNNDVYGLPKPFDVTLVTEGNAKNGKP NKTSDAPGSPGSGSGSGSSTKSSNVKLIIVPVGGVAGCLVGLGVCILYAKK RKPARVQSPGSNNVHPHSGNDNDIKLTVASSLNSGGSGSGLSGSAAADIHVV EAGNLVITLVLRNVNTNFEENILGRGFTVYKGLHDGTIKIAVKRMSSVVSQK LTFPKLSITVLTKMRHRLVALLGYCLDGNRLLVYEMPQGTLSQHLFWKEGRKP LDWTRRLAIDVARGVYHLTLAHQSFIVHDLKPSNILLGDDMKRAKVSDFGLVRLAP DGKYSIETRVAGTGYLAPYAVTGVTTKVDIFSLGVILLMELITGCKALDETQPEDS VHLVWTRRVAASNDENAFKNAIDPNISLDDDTVASIEKWLGHACCAREPTQPEDM AHVKNQTSISITRPSGFADSPFTSDGR"
 complement(7541. .7564)
 /rpt_family="AT_rich"
 complement(9164..13791)
 /gene="AT2g01830"
 /note="synonym: T23K3.2; supported by cDNA:"

mRNA
 CDS
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 gene
 mRNA
 CDS
 repeat_region
 gene

gi_13359154 dbj AB049934.1 AB049934.1
 complement(join(<9164. .9708,9850. .11054,11138. .11236, 11315. .11544,11638. .11876,11958. .12176,12352. .12521, 12615. .12851,13032. .13581,13699. .13791))
 /gene="AT2g01830"
 complement(join(9477. .9708,9850. .11054,11138. .11236, 11315. .11544,11638. .11876,11958. .12176,12352. .12521, 12615. .12851,13032. .13574))
 /gene="AT2g01830"
 /codon_start=1
 /product="putative histidine kinase"
 /protein_id="AAD21777.2"
 /db_xref="GI:20198055"
 /translation="MNAALNNHDEEPRRIEISDESLENLKSDFYQGGGALN SSKPRKIDFMRSGFAKQOQQLQHSVAVKMNNDLNGKKGSTFIOEHRA LKALILWIIIVGFTSSGIYQWMDANKIRBEVLVMSQDQARMKQDQSVSVNHV HALAILVTFHYHKNPSAIQDETFAEYTAFTAPRLPSGVAYAKVNVNPERFERQ HNWIKTMDGEPSVURDEVAPVIFSQDSVYLSLMSGEEDENILRARETKAV LTPFRLLETHLGVYDITNASDPLVYKSSLPENFTVEERIAATAGYLGGADEVSLNLLG QLAGNQAIVVHYDITNASDPLVYKSSLPENFTVEERIAATAGYLGGADEVSLNLLG QKAPILNVLTVPLFAIGLVGYLYLGAAMHIVKVEDDFHEMOELKLVRAAADVAK SQPLATVSHERTPMGILGMLAMLLDTLSTORDVATQVCGKALIALINEVLDL RQIILNVCNSVKTSEKHFVKVHLAEOQSDSEKPNALNGVSEEMIVVSKOSSYN KTEAGKLELSEVPEDIRSLDDVLSLFSSESNNKSTELAVFVSDKVEIIVKSGSRF MPEQADSTSRNYGGTGIGLSIKCLVELMRGOINFIRPHIGSTFWTAVLEKCDK CSAINMKKENVEHLPSTFKGMKAIIVDAKVRAAVTRHYMKRLGINDVNVVTSKTAV VAAATNNSFDRKAGPADVTIMKPLRASMITGSDNDSIEIRLLNSRTNGVHHKSPKLA LPATNITNSFDRKAGPADVTIMKPLRASMITGSDNDSIEIRLLNSRTNGVHHKSPKLA LKSLTGKTLVVDNINVRRAAGALKKFGAASVCAESQVAGLGLQIHPHTFACPM DIMPQNDGFETRAQRMMEKTKETKNTLEWHLPILANTADVIHATIEECLKSGNDGY VSKPFEENLYSKVAKSPKPNPISPS"
 14727..14747
 /rpt_family="AT_rich"
 16299..16340
 /rpt_family="AT_rich"
 complement(16407..16444)
 /rpt_family="TAAAA)n"
 complement(16445..16471)
 /rpt_family="AT_rich"
 complement(18131..18161)
 /rpt_family="AT_rich"
 complement(18930..18952)
 /rpt_family="AT_rich"
 complement(19270..24592)
 /gene="AT2g01840"
 /note="synonym: T23K3.3"
 complement(join(<19270. .20837,20891. .23417,23540. .>24592))
 /gene="AT2g01840"
 complement(join(19270. .20837,20891. .23417,23540. .24592))
 /gene="AT2g01840"
 /codon_start=1
 /product="putative non-LTR retroelement reverse transcriptase"
 /protein_id="AAD21778.1"
 /db_xref="GI:4522005"
 /translation="MADNRRALQNLINLGIDETVALPQAIIVQRAVDNDFCLIGRPL MPNQLRQILTTVPTWGLVFGVRGRIQNRFPQIFPSEESLDTVLRRGMSFADR MLVVEWTPMDPLVNFIPFMIQVREIPLQFLNLEVIDNIAGSLGRKAVDFDPFTT TRVEFQIQIKWDVNPFLPQRNYQFSLGVNTVLSFYERLGRFCDCVGRMTHDAGSC VLQVGRALDGDGDDNADQGFPHGNPVIQIIGDDQPIQDGEVEEIVPQVGA DYNAKEEVSIDIDPHTNALEDISSDNRNMFHEWTCDEDLVYNPVPSPANATGDIILG PVLLSQELRKRRHNFEESGAGTTSPMRVFWNCQGLQPLTVRLREVSQVRYFLD MLPLIKQDQNVTRDLGVKMGPEDCIISPRGLSGVLVYVKKHLSIQVISHDRLV DLVVEKNFNYLSCLYGHPIPSERHLMELQVRSNHRSGPMWCGDNEILNLNEK KGRRASISLQNTMNCNNMKDLKSGNPYSWVGKQNETIESCLDRVFNDSHQ ASPAPTEFLPIAGSHAPVIIDIAEEVCTKRGQFRYDRHFQDFEDFVDSVQGMNR LQDLNQAQVREDELYWHLKSRNRMLLGDRNTFFVASTLKRKSRNRIKAITDOAQGE NRDDTIGKVAENVFADLTFTTQTSDEWEIISGAPKPTQMNHLLQSVTDOEVRD VFAIGADRAGFDGFTAAFYHLDLINDVCLMVHFFPESDYMNDQINQTOICLIPK IIDPKHMSDTRISLCTCTATKISKILIKKLOGLDGVSDSQAFVPGNQISDNVLV ABELHLSLRSRRCQSGYAVKTDISKAYDRVENVFLEKVMYIQLGFAPRVWKMTCV

```

TSVSYEVNLGSPYKIFPSRGIHQDPLSPFLFCAEVLNMLKAEVNLQIHCMK
ITKDCIAIHLFLPADSLFFCRASNQNIQALIFKKYBEASGCKINYAKSSIIFFQK
IPTMRQRIHLGIDNVRGGKYLGLPEQLGRKVELEFYIVTKVKTGWAYNYL
SPAGKEIVITKAMALPVYSMNCFLPTLINCINSLITAFWKGKNEGDLGFKOLHQ
FNRLAKQAWRLITNPQSLARLYKGLYPNTYLRANKGGHASYGWNISIOEGKLLL
QOGLVRLGDCQTKIWEDEPWLPTLPPRARGPILDEMKVADLRKREKDEPVILE
GVLNPDQQLAKSLYLSUNSAARSKWATYRNTQYTVRSGYVWATHVTEEBEINPL
EGDVLKQIRWLKIPKIKHFTWRCLSGALSTTQLRNRNIPADPTQRCNCNABETI
NHIIFTCYQVWNSANFSNGRLCFTDNLENIRLIILQKKKNQNLPTLGLMPPFI
MWRLKSREYILFQDLRFPWKVAQEQEAVETWMTWNTDIAISHNTAQSDNRPISR
SKWMSPPRGFLKCNFDSGVVQGRDVTSGWILRDCNGRVLHSGCAKLOQSYALQAE
ALGFLHALQWIRGYCYWFEEDNLELNLINKTDEHLLTLLYDIRFWTKLPFS
SIGVYNRNLRAADKUTKYSANSSLYTFHVPWRQLYLYIFFTN"
repeat_region complement(24924..24969)
/rpt_family="AT_rich"
25027..25048
/rpt_family="AT_rich"
/rpt_family="AT_rich"
complement(25518..25541)
/rpt_family="AT_rich"
complement(28275..28319)
/rpt_family="AT_rich"
complement(28625..28645)
/rpt_family="(A)n"
complement(30034..30080)
/rpt_family="AT_rich"
30639..30666
/rpt_family="AT_rich"
31150..31187
/rpt_family="AT_rich"
31368..31364
/gene="At2g01850"
/notes="synonym: T23K3.4; identical to endoxyloglucan
transferase related protein EXGT A3 GI:2154609, GI:5533313
from (Arabidopsis thaliana); supported by full length cDNA:
Cares-18876"
join(31368..31702,31777..31877,32003..32211,32776..33464)
/gene="At2g01850"
join(31519..31702,31777..31877,32003..32211,32776..33283)
/gene="At2g01850"
/codon_start=1
/product="xyloglucan endotransglycosylase (EXGT-A3)"
/protein_id="AA021783.1"
Query Match 8.7%; Score 43.8; DB 8; Length 73009;
Best Local Similarity 51.3%; Pred. No. 0.93;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 281 CGAAGATGCGAAACTTGGCAGTTTTCAGTCTTTTCAGTAAAGTGGCTTTTTCACCTCC 340
DB 24845 CGAAATCGCCTTAACAATGCCATTTTGTGTTTCTGAGAGTCTTTTAT 24904
QY 341 ACAACTTATTTAGCCTTGATATGCTTATATAGCTGATGCACTGATGTTGTATACC 400
DB 24905 CTAGCTTTTGTATCTCTATTTTATTAAGAATTTTCTTATTAATGATTAAT 24964
QY 401 TGAATTTCTGTACAGCAAAATGTGAGAGTTTATTTCTTCACTATATCATCCGTTTAA 460
DB 24965 ATTACTATCACCTAGTAAATAAAATAAATCTCTAGAAACAATAGCACTCTTTGAA 25024
QY 461 GCATATTTATTTCTTATA 479
DB 25025 ACATTTTATTTTATTTT 25043
RESULT 9
AC102623/c
LOCUS AC102623 259375 bp DNA linear HTG 26-MAR-2004
DEFINITION Mus musculus chromosome 9 clone RP23-441D19 map 9, *** SEQUENCING
IN PROGRESS ***, 10 unordered pieces.
ACCESSION AC102623
VERSION AC102623.5 GI:45752824
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)

```

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 259375)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 9, clone RP23-441D19

Unpublished

2 (bases 1 to 259375)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Chospel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAngelis, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 259375)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chospel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeAngelis, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (26-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 26, 2004 this sequence version replaced gi:28927714.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submission@broad.mit.edu

----- Project Information

Center project name: L19123

Center clone name: 441_D_19

* NOTE: This is a 'working draft' sequence. It currently

* consists of 10 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 40149: contig of 40149 bp in length
40150: gap of 100 bp
40249: contig of 9846 bp in length
50095: gap of 100 bp
50195: gap of 100 bp
50196: contig of 45443 bp in length
95638: gap of 100 bp
95639: gap of 100 bp
95738: gap of 100 bp
107514: contig of 11776 bp in length
107614: gap of 100 bp
107615: contig of 30282 bp in length
137897: gap of 100 bp
137898: contig of 30282 bp in length
137996: gap of 100 bp
140291: contig of 2295 bp in length
140292: gap of 100 bp
140391: gap of 100 bp
140392: contig of 19655 bp in length
160047: gap of 100 bp
160146: gap of 100 bp
160147: contig of 32380 bp in length
192526: gap of 100 bp
192527: contig of 32380 bp in length
192626: gap of 100 bp
192627: contig of 48548 bp in length
241174: gap of 100 bp
241175: gap of 100 bp
241274: gap of 100 bp
259375: contig of 18101 bp in length.

FEATURES
source
1. 259375
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP23-441D19"
/clone_lib="RPCI-23 Female Mouse BAC"

ORIGIN

Query Match 8.7%; Score 43.8; DB 2; Length 259375;
Best Local Similarity 47.6%; Pred. No. 0.79;
Matches 129; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 84 GGAGTACACCAAGTATACATCCAGTCATCATGTAATCTTCTGGTACAGATGTAC 143
DB 201975 GGAGATTCATATGATGCTGATGTAATCTTCACTGCGTCCCTGGAAGGCAC 201916
QY 144 ATTGGATGGAATGATGCTTTTGGCTTGATTTCTATGGCGAGCATATCTCTGGATCCC 203
DB 201915 ATTATATACAGACAGAACTTTGAGTATATCTCTGGTGGATGATATTTTACTT 201856
QY 204 AGAGGATATGTTAGTATGTAAGATTTGGGGGAAATACCGGATGAATACAGAGGAGCT 263
DB 201855 TGTGTTGATGCTAATGATATCTTCTCGGAAAAATACAGCTGCATTACATCTTAAGCA 201796
QY 264 CCACATGTTTTTAAGCGGAAGATGGCAAACTTTGCGAAGTTTTCATCTTTCAGTAAG 323
DB 201795 ACATAAATCTGAGGCGCACTTAATGTTTACATGATGATGCTTTTGTCTTCC 201736
QY 324 TGGCTTTTTCACCTCCACAACTATTTTAG 354
DB 201735 TGGTCTTTTCAGCAGAGTGAATTTCTAG 201705

RESULT 10
BX248091/c
LOCUS BX248091 165075 bp DNA linear VRT 17-JUN-2004
DEFINITION Zebrafish DNA sequence from clone DKEY-25719, complete sequence.
ACCESSION BX248091
VERSION BX248091.12 GI:48675761
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 165075)

Chapman, J.

Direct Submission

Submitted (17-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jun 12, 2004 this sequence version replaced gi:4776069.

COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep/Clone-derived
zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-25719
is from a zebrafish BAC library

VECTOR: pindigobAC-5.

FEATURES

source

1. 165075
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-25719"
/clone_lib="DanioKey"

ORIGIN

Query Match 8.5%; Score 42.6; DB 5; Length 165075;
Best Local Similarity 51.9%; Pred. No. 1.7;
Matches 96; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 306 TTTGATCTTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTATTTAGCTTGCATATG 365
DB 101727 TTAGCATTTTCCATATTTGCCAAATTTCTCTCAGCCAGCTTTTGTCTTACCTAAG 101668
QY 366 CTTATATATAGCTGATTGCAACTGTAGTTGTACCTGATTTCTCTGTACAGCAAAATGTG 425
DB 101667 CAAATATTAATTTGATTTGATGATGTTTTTAAATATGCTTTTTTAAATCTTAAGTAAAG 101608
QY 426 AGAGTTTATTTCTTCAACTATATCCATCCGTTTAAGCATATTTTATTTCTTATCTGGC 485
DB 101607 ATGGATTTAATACCCCAAAAACCTAGTCATTTTAAACCAACTTTGATTTCTTATTTGTG 101548
QY 486 TTCGT 490
DB 101547 ATTGT 101543

sequence.
 AL591624
 VERSION AL591624.6 GI:14456008
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 146887)
 Tracey, A.
 Direct Submission
 Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 14, 2001 this sequence version replaced gi:14280475.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; Sw.,
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/c_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-143A22 is from the library RPCT-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6
 This sequence is the entire insert of clone RP11-143A22.
 Location/Qualifiers
 1. 146887
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone_lib="RPCT-11.1"
 /clone="RP11-143A22"
 1267..1685
 /note="MSTA repeat: matches 3..426 of consensus"
 2139..2243
 /note="MIR repeat: matches 65..179 of consensus"
 3076..3971
 /note="LIPB4 repeat: matches 5290..6182 of consensus"
 4145..4226
 /note="41 copies 2 mer tt 63% conserved"
 4727..5146
 /note="MSTB repeat: matches 1..426 of consensus"
 5248..5390
 /note="MLTIB repeat: matches 404..547 of consensus"
 5594..5668
 /note="MLTIIH repeat: matches 72..156 of consensus"
 7466..7754
 /note="LIPPA repeat: matches 5859..6163 of consensus"
 7770..8441
 /note="LIPPA repeat: matches 5178..5854 of consensus"
 9962..10091
 /note="MIR repeat: matches 89..249 of consensus"
 10910..10971
 /note="31 copies 2 mer at 72% conserved"
 11187..11254

repeat_region
 /note="34 copies 2 mer tt 69% conserved"
 11440..11460
 /note="LI repeat: matches 5257..5277 of consensus"
 11461..12131
 /note="LIMA3 repeat: matches 5568..6304 of consensus"
 12132..12381
 /note="ALUSC repeat: matches 56..305 of consensus"
 12382..12548
 /note="LIMA3 repeat: matches 5413..5568 of consensus"
 12549..14668
 /note="LI repeat: matches 3127..5257 of consensus"
 14669..14934
 /note="ALUJO repeat: matches 23..283 of consensus"
 14935..15490
 /note="LI repeat: matches 2594..3127 of consensus"
 15491..16523
 /note="MER11C repeat: matches 1..1071 of consensus"
 16524..16707
 /note="LI repeat: matches 2410..2594 of consensus"
 16707..19324
 /note="LIMI repeat: matches -1208..1472 of consensus"
 19325..19465
 /note="ALUJB repeat: matches 1..142 of consensus"
 19498..20080
 /note="LIR9 repeat: matches 14..625 of consensus"
 20083..20154
 /note="ALUJ/FRAM repeat: matches 223..294 of consensus"
 20198..20370
 /note="LIMI repeat: matches -1388..-1207 of consensus"
 22293..22436
 /note="MIR repeat: matches 6..154 of consensus"
 22901..23283
 /note="MLTIB repeat: matches 1..390 of consensus"
 24998..25313
 /note="ALUJB repeat: matches 1..306 of consensus"
 26067..26223
 /note="MIR repeat: matches 2..155 of consensus"
 26239..26573
 /note="LIPAS repeat: matches 5536..5873 of consensus"
 26574..26850
 /note="LIPAS repeat: matches 5866..6142 of consensus"
 26872..26927
 /note="MIR repeat: matches 196..252 of consensus"
 27819..28128
 /note="ALUSQ repeat: matches 1..307 of consensus"
 31725..31756
 /note="16 copies 2 mer ac 90% conserved"
 31769..32127
 /note="MLTIIA repeat: matches 1..363 of consensus"
 33434..33481
 /note="24 copies 2 mer at 75% conserved"
 33547..35944
 /note="TIGGER1 repeat: matches 11..2418 of consensus"
 37127..37298
 /note="MIR repeat: matches 12..192 of consensus"
 37299..37476
 /note="ALUJO repeat: matches 122..301 of consensus"
 37623..37932
 /note="ALUSQ repeat: matches 1..310 of consensus"
 40111..40432
 /note="161 copies 2 mer tt 54% conserved"
 40680..40797
 /note="MIR repeat: matches 48..191 of consensus"
 43332..44001
 /note="MER44C repeat: matches 31..728 of consensus"
 44203..44671
 /note="MLTID repeat: matches 1..505 of consensus"
 45722..46278
 /note="L2 repeat: matches 2143..2710 of consensus"
 46501..46665
 /note="LIMB1 repeat: matches 6002..6168 of consensus"
 46686..46792
 /note="L2 repeat: matches 2053..2163 of consensus"


```

LOCUS      AC067897      178951 bp      DNA      linear      HTG 22-JUL-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-199A24 map 6, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION  AC067897
VERSION     AC067897.3 GI:9369516
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 178951)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagoo,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,D., Karatas,A.,
            Klei,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
            Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainou,J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 22, 2000 this sequence version replaced gi:8746751.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L7658
            Center clone name: 199_A_24
            ----- Summary Statistics
            Sequencing vector: M13; M7815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 171035 bases at least Q40
            Consensus quality: 175159 bases at least Q30
            Consensus quality: 176736 bases at least Q20
            Insert size: 181000; agarose-fp
            Insert size: 177551; sum-of-contigs
            Quality coverage: 4.7 in Q20 bases; agarose-fp
            Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 1324: contig of 1324 bp in length
            * 1325 1424: gap of 100 bp

FEATURES             source
                     1..178951
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="6"
                        /clone="RP11-199A24"
                        /clone_lib="RPC1-11 Human Male BAC"
                     1..1324
                        /note="assembly_fragment"
                        1425..3037
                        /note="assembly_fragment"
                        3138..5340
                        /note="assembly_fragment"
                        5441..10028
                        /note="assembly_fragment"
                        10129..15006
                        /note="assembly_fragment"
                        15107..20166
                        /note="assembly_fragment"
                        20267..28179
                        /note="assembly_fragment"
                        28280..37701
                        /note="assembly_fragment"
                        37802..51148
                        /note="assembly_fragment"
                        51249..65912
                        /note="assembly_fragment"
                        66013..82263
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:left"
                        82364..99075
                        /note="assembly_fragment"
                        99176..120890
                        /note="assembly_fragment"
                        120991..147381
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:right"
                        147482..178951
                        /note="assembly_fragment"

ORIGIN
Query Match      8.3%; Score 41.8; DB 2; Length 178951;

```

```

Best Local Similarity 53.3%; Pred. No. 2.7;
Matches 88; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 314 TTTCAGTAAGTGGCGCTTTTTCACCTCCACCACTTATTTTATAGCCCTTGATGCTATATATA 373
    |||||
Db 166854 TTTCAGCCATGACCTTATTTTCAACACCTCAGGATAAGCTAGCAGCCACATTTTAAATA 166795
    |||||

Qy 374 TAGCTGATGCACTGTAGTGTGTACCTGATTCCTGTTACAGCAATGTGAGAGTTTT 433
    |||||
Db 166794 AAGATGTTCTCATGTGCTGATGTTTAAATATTTTTCAGCTAGGAGAGCTTAAATGTA 166735
    |||||

Qy 434 ATTCTTCAACTATATCATCGTTTAAAGCATATTTTATTTCTTAT 478
    |||||
Db 166734 GTTTTAAAGATATCTTGTGTTGAGAGATTATCTCCCTTTT 166690
    |||||

RESULT 15
AL365315/c
LOCUS
DEFINITION Human DNA sequence from clone RP11-275K8 on chromosome 1 Contains
the gene for NY-REN-45 antigen, the 3' end of a novel gene and a
CpG island, complete sequence.
ACCESSION AL365315
VERSION 8 GI:10120104
KEYWORDS HTG; CpG island; NY-REN-45.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 139118)
Wallis, J.
Direct Submission
Submitted (07-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Sep 13, 2000 this sequence version replaced gi:9994248.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
This sequence is the entire insert of clone RP11-275K8. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-275K8 is from
the library RPCI-11.1 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES
Source
1..139118
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clones="RP11-275K8"
/clone_lib="RPCI-11.1"
928..1232
/notes="MER49 repeat: matches 5. .318 of consensus"

repeat_region
/notes="MER49 repeat: matches 318. .923 of consensus"
complement(4382..5034)
/notes="match: GSS: Em:AQ282711"
match: STS: Em:G55575"
complement(6338..6739)
/notes="match: GSS: Em:AQ112146"
complement(7835..8439)
/notes="match: GSS: Em:AQ061503"
10093..10339
/notes="match: GSS: Em:AQ201365"
complement(12776..13192)
/notes="match: GSS: Em:AQ0900482"
complement(12829..13119)
/notes="match: GSS: Em:AQ091519"
complement(12896..13424)
/notes="match: GSS: Em:AQ498487"
complement(12942..13426)
/notes="match: GSS: Em:AQ768825"
13495..13644
/notes="MER51B repeat: matches 449. .600 of consensus"
13655..13916
/notes="L1PA5 repeat: matches 11. .6145 of consensus"
20389..20757
/notes="MER51-internal repeat: matches 1166. .1513 of
consensus"
21021..21239
/notes="MER83-internal repeat: matches 1652. .1858 of
consensus"
complement(21144..21575)
/notes="match: GSS: Em:AQ193572"
25208..25346
/notes="LTR41 repeat: matches 3. .141 of consensus"
25347..25609
/notes="L1MEC repeat: matches 2390. .2303 of consensus"
28928..28959
/notes="8 copies 4 mer gtgt 93% conserved"
31621..33230
/notes="L1MA4A repeat: matches 4674. .6296 of consensus"
33521..33989
/notes="L1MA4A repeat: matches 4203. .4674 of consensus"
34298..37553
/notes="L1MA4A repeat: matches 920. .4203 of consensus"
37782..39363
/notes="L1MA4A repeat: matches -679. .920 of consensus"
39430..39775
/notes="L1MEC repeat: matches 3931. .4269 of consensus"
40070..41707
/notes="L1MEC repeat: matches 2323. .3931 of consensus"
42020..42038
/notes="L1MEC repeat: matches 2559. .2323 of consensus"
42444..42651
/notes="L1MEC repeat: matches 2357. .2560 of consensus"
45347..45392
/notes="MUR2B repeat: matches 395. .440 of consensus"
47271..47476
/notes="MER30 repeat: matches 4. .230 of consensus"
complement(47316..47757)
/notes="match: GSS: Em:AQ602835"
48619..49377
/notes="CpG island"
/evidence="not experimental"
48839..103260
/genes="BA275K8.1"
join(48839..49222,55240..55293,55527..55572,57355..57428,
59124..59182,59455..59535,60454..60591,61363..61453,
67949..68139,76809..76924,83320..83407,83538..83654,
85585..85755,89470..89625,93278..93374,100339..100523,
100606..100744,101510..103260)
/genes="BA275K8.1"
/product="BA275K8.1 (NY-REN-45 antigen)"
/notes="match: CDNAB: Em:AF155110 Em:AY007117
match: EST6: Em:AI540877 Em:AI394692 Em:AA415243 Em:W07820"

```

```

Em:AI548144 Em:AI906920 Em:AA188593 Em:R13443 Em:AA159862
Em:AI904415 Em:Z41959 Em:AW503472 Em:AA433833 Em:AI984556
Em:AI686848 Em:AA936410 Em:T36035 Em:T31823 Em:T31293
Em:AA505490 Em:AI339212 Em:AA654834 Em:D20344 Em:AA993194
Em:AI347931 Em:AI765832 Em:AW264231 Em:AW193971
Em:AW195137 Em:N56164 Em:AW275527 Em:N29314 Em:W00437
Em:W92917 Em:T31790 Em:W81046 Em:AA025593 Em:Z43271
Em:AA119267 Em:AA780119 Em:AA249138 Em:AI907040
Em:AA858164 Em:AI334636 Em:W58357 Em:AI914137 Em:AI185187
Em:AI393364 Em:AW264326 Em:D57384 Em:Z24908 Em:N41386
Em:H15307 Em:AA740162 Em:N98650 Em:N72757 Em:T36257
Em:AW572796 Em:AA371494 Em:AA480197 Em:AA970708 Em:W81099
Em:AI308168 Em:H51146 Em:N58984 Em:AL048814 Em:AI992162
Em:T30120 Em:AW005644"
/evidence=not experimental
join(49140..49222,55240..55293,55527..55572,57355..57428,
59124..59182,59455..59535,60454..60591,61363..61453,
67949..68139,76809..76924,83320..83407,83538..83654,
85585..85755,89470..89625,93278..93374,100339..100523,
100606..100744,101510..102071)
/genes="ba275k8.1"
/note="match: proteins: Tr:Q62523 Tr:Q9Y597"
/codon_start=1
/evidence=not experimental
/product="ba275k8.1 (NY-REN-45 antigen)"
/protein_id="CAD20310.1"
/db_xref="GI:18093151"
/db_xref="GOA:Q8WX16"
/db_xref="UniProt/TREMBL:Q8WX16"
/translation="MAGHGCGFPAAAGSGEIVQNVGTFPSTRSQTLMPIDSPFF
SLLSGRIETLRDETGAIFIDRDPAPAFILNFKELDLRGVSINVLRHEAEFYGI
TFLVRLLELCEELERSGCVLFGVLPPLPGIPSRKINNTVRSADNRGLNSTEGEAR
GNGTQVLSTGTEVRLGPDPVKVLIVAGHNMIVAAAHFAVCYRIKESGWOQ
VFTSPVLDLTERVALNAKVGGPHGDKMVAASESIIILWSVDGSGSGEIGVFS
LGVVDALFFIGNQLVATSHGKGVGNVAOTQHWQVDVPITSDTAGSPILLGCNN
GSYYIDMQKFLPRLMNDLLVELYDPSNDAITALSVLTPKTSVSGNWIEIAYGT
SSGAVRIVQHPETVSGSPQLQFTFVHRSPVTKIMLSEKHLVSCADNNHWTMTVT
RFRGMISTQPGSTPLASFKILSLEETESHGYSVSGNDIGPFGERDDQVFIQKVPIT
NKLTVRLSTGKRICIOAVDCTTISFTVRECEGSRMGSRPRRYLFTGHTNGSIQM
WDLTTAMDVNKSEDKDVGPTTEELIKLLDODCLSTRCATNPISPAISVVOHSHLR
ENSSQLQHDHDTHEAATYGSNRPYRESPLARARTESFHSYRDFQINLRNVER
AVPENENLGPQAEVKGATGECNISKSPGVEIKSRLDSDGLEVHKIAEGRFSESK
RSEDENENKIEFRKKGFGGGFLGRKKVPLASSPSTSDGDTSPGTASPSPTKT
PSPRHKKSDSGQEYSL"
complement(51320..51833)
/note="match: GSS: Em:AQ706317"
complement(51368..51821)
/note="match: GSS: Em:AQ572973"
complement(52522..52540)
/note="Single clone region. Restriction digest data
(BAMHI, ECORI and HINDIII) confirms assembly along with
pUC read pairs."
complement(54241..54435)
/note="match: GSS: Em:AQ111706"
60256..60317
/note="31 copies 2 mer tt 69% conserved"
62961..63384
/genes="ba275k8.1"
/note="match: GSS: Em:AQ771889"
65118..65133
/note="HERVK4 repeat: matches 5661..5682 of consensus"
65134..66431
/note="MER41-internal repeat: matches 2662..3944 of
consensus"
66427..66511
/note="MER57-internal repeat: matches 87..875 of
consensus"
67834..67875
/note="21 copies 2 mer tt 81% conserved"
71022..71080
/note="MER68 repeat: matches 353..411 of consensus"
72704..72799

```

```

Query Match      8.3%; Score 41.6; DB 9; Length 139118;
Best Local Similarity 62.5%; Pred. No. 3.1;
Matches 65; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 368 TATATATAGCTGATTCGCACTGCTAGTTGTTACCTGATTTCCCTGTACACGCCAAATGAG 427
DB 9590 TAGATGTGACTGATTCGCCAAAGTTTCAGCTTCCATATTCGCTGTAGTCCACAGTGCT 9531

QY 428 AGTTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTAT 471
DB 9530 TGTACATTTATTAATAATACACATACGTACATGATAATTAAT 9487

```

Search completed: July 12, 2005, 05:59:25
Job time : 2543.52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 382.755 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929C-1_COPY_4900_5400

Perfect score: 501

Sequence: 1 aatagtttaattctgtcc.....tggtcttgtaaccatgcac 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10329	4	AAA89385 Arabidops
2	501	100.0	10329	4	AAA89385 Arabidops
3	244.6	48.8	6571	4	AAA89354 Arabidops
4	60	12.0	467	4	AAA89379 Brassica
5	43.8	8.7	70768	6	AAL41152 Wooden le
C 6	40.2	8.0	3357	6	ABN70900 Streptoco
C 7	40.2	8.0	3537	6	ABN66624 Streptoco
C 8	40.2	8.0	110000	6	ABN71527_06 Continuation (7 of
9	39.2	7.8	9728	6	ABK40092 Human che
10	39.2	7.8	9728	6	ABL33903 Human imm
11	38.2	7.6	18585	6	ABL34608 Human met
12	38.2	7.6	18585	7	ADS99869 Bisulphit
C 13	38	7.6	6283	6	ABL32089 Human imm
C 14	37.4	7.5	3987	10	ADC99111 Human KPP
C 15	37.4	7.5	4613	10	ADH29035 Human chr
C 16	37.4	7.5	4616	13	ACN39532 Tumour-as
C 17	37.4	7.5	4855	2	AAV22252 Human myo
C 18	37.2	7.4	2414	4	ABL27250 Drosophil
19	37.2	7.4	46819	9	ACD19059 E. coli 0
20	37.2	7.4	46897	10	ADC00585 Enterohae

C 21	37.2	7.4	78082	12	ADQ97968 Human can
C 22	37	7.4	671	5	ADI67504 Human ova
C 23	37	7.4	671	5	ADI73894 Human ova
C 24	36.8	7.3	701	4	AAD17783 Zea mays
C 25	36.8	7.3	1204	4	AAD17781 Zea mays
C 26	36.8	7.3	2000	8	ADA71938 Rice gene
C 27	36.8	7.3	15620	2	AAV03553 DNA seque
C 28	36.8	7.3	49914	3	AAH81471 N. mening
C 29	36.8	7.3	110000	3	AAH81490_05 Continuation (6 of
C 30	36.8	7.3	349980	3	AAF21607 Neisseria
C 31	36.2	7.2	127369	11	ACM44006 Human gen
C 32	36	7.2	2526	12	ADI42690 Plant tra
C 33	36	7.2	2526	12	ADO29339 Corn orth
C 34	35.8	7.1	96596	9	ADA02930 Human PIK
C 35	35.8	7.1	96596	10	ADB72668 Human PIK
C 36	35.8	7.1	96596	10	ADC85410 Human car
C 37	35.8	7.1	96596	12	ADM74525 Human car
C 38	35.8	7.1	175590	10	ADD50650 BAC seque
C 39	35.6	7.1	1548	10	ACF67840 Photorhab
C 40	35.6	7.1	60203	10	ACF65383_4 Continuation (5 of
C 41	35.6	7.1	110000	10	ACF67367_04 Continuation (5 of
C 42	35.6	7.1	110000	10	ACF67367_05 Continuation (6 of
C 43	35.4	7.1	22930	4	AAK78545 Human imm
C 44	35.4	7.1	138363	13	ABD32624 Human can
C 45	35.2	7.0	2000	8	ADA71938 Rice gene

ALIGNMENTS

RESULT 1
AAA89385
ID AAA89385 standard; DNA; 10329 BP.
XX
AC AAA89385;
XX
DT 04-NOV-2004 (revised)
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT exon 1..1008
FT /*tag= a
FT /*number= 1
FT intron 1009..1295
FT /*tag= b
FT /*number= 1
FT exon 1296..2550
FT /*tag= c
FT /*number= 2
FT CDS 1301..9808
FT /*tag= d
FT /*product= "Columbia protein"
FT /*note= "This sequence contains 15 introns."
FT intron 2551..2673
FT /*tag= e
FT /*number= 2
FT exon 2674..2752
FT /*tag= f
FT /*number= 3
FT intron 2753..2867
FT /*tag= g
FT /*number= 3
FT exon 2868..3113
FT /*tag= h
FT /*number= 4
FT intron 3114..3506
FT /*tag= i

FT FT /number= 4
FT exon 3507. .3680
FT /tag= j
FT intron /number= 5
FT 3681. .3973
FT /tag= k
FT exon /number= 5
FT 3974. .4895
FT /tag= l
FT /number= 6
FT allele 4338
FT /tag= m
FT /note= "ecotype Zurich"
FT intron 4896. .4975
FT /tag= n
FT exon /number= 6
FT 4976. .5217
FT /tag= o
FT intron /number= 7
FT 5218. .5777
FT /tag= p
FT exon /number= 7
FT 5778. .5882
FT /tag= q
FT /number= 8
FT intron 5883. .6082
FT /tag= r
FT exon /number= 8
FT 6083. .7480
FT /tag= s
FT /number= 9
FT allele 6721
FT /tag= t
FT /note= "ecotype Zurich"
FT intron 7481. .7615
FT /tag= u
FT exon /number= 9
FT 7616. .7771
FT /tag= v
FT intron /number= 10
FT 7772. .7914
FT /tag= w
FT /number= 10
FT exon 7915. .8070
FT /tag= x
FT /number= 11
FT intron 8071. .8153
FT /tag= y
FT /number= 11
FT exon 8154. .8318
FT /tag= z
FT intron /number= 12
FT 8319. .8451
FT /tag= aa
FT exon /number= 12
FT 8452. .8629
FT /tag= ab
FT /number= 13
FT intron 8630. .8718
FT /tag= ac
FT /number= 13
FT exon 8719. .8918
FT /tag= ad
FT /number= 14
FT intron 8919. .9000
FT /tag= ae
FT /number= 14
FT exon 9001. .9211
FT /tag= af
FT /number= 15
FT intron 9212. .9284
FT /tag= ag
FT /number= 15

FT exon 9285. .10329
FT /tag= ah
FT /number= 16
FT XX
FT PN WO200100801-A2.
FT XX
FT PD 04-JAN-2001.
FT XX
FT PF 21-JUN-2000; 2000WO-EP005761.
FT XX
FT PR 23-JUN-1999; 99GB-00014623.
FT XX
FT PA (NOVS) NOVARTIS AG.
FT PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT XX
FT PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
FT XX
FT XX WPI; 2001-137952/14.
FT DR P-PSDB; ABE20063.
FT XX
FT PT Novel gene encoding a protein that controls gene silencing, in particular
FT PT silencing of plant genes.
FT XX
FT PS Claim 5; Page; 48pp; English.
FT XX
FT CC The present sequence is that of an Arabidopsis thaliana ecotype Columbia
FT CC gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino
FT CC acid protein with similarity with ATPase/helicase proteins of the
FT CC SWI2/SNF2 family, and which controls gene silencing, particularly plant
FT CC gene silencing. Trans-acting modified loci were identified by T-DNA
FT CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
FT CC carries a heritable inactivated, methylated hygromycin resistance gene
FT CC (hyg). Mutation of a silencing modified locus results in release of
FT CC silencing of the hyg gene and restores hygromycin resistance. Plants
FT CC homozygous for the silent resistance gene were subjected to
FT CC transformation with a bar selectable marker gene under control of the T-
FT CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
FT CC screened for hygromycin resistance. The mutant phenotype (hygromycin
FT CC resistance) was screened for genetic co-segregation with a specific T-DNA
FT CC insert. Cloning of the tagged gene allowed characterization of the
FT CC present sequence. 2 Allelic differences were found in the corresponding
FT CC gene of A. thaliana ecotype Zurich, occurring at nucleotide positions
FT CC 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful
FT CC as a molecular tool for regulating gene expression. Note: the present
FT CC sequence is not shown in the specification but is derived from the A.
FT CC thaliana mutant silencer gene sequence given in AAA89353
FT CC
FT CC Revised record issued on 04-NOV-2004 : Correction to feature table key
FT XX
FT SQ Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;
FT
FT Query Match 100.0%; Score 501; DB 4; Length 10329;
FT Best Local Similarity 100.0%; Pred. No. 8.9e-133;
FT Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 1 AATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTACTTTGCTG 60
FT
FT Db 4900 AATAGTTTAATTTCTGCTCCGATACCTCTAGTGTTCATTGATTATGCAACTACTTTGCTG 4959
FT
FT QY 61 ACTATCTTCTCCTACAGGGGCTTTGAGTCACCAAAAGTCATAAAGTCAGTCATCATGT 120
FT
FT Db 4960 ACTATCTTCTCCTACAGGGGCTTTGAGTCACCAAAAGTCATAAAGTCAGTCATCATGT 5019
FT
FT QY 121 AAATCTTCTGCTACAGATGGTACATTGGATGGAATGATGCTTTTGGCTTGTATTCTATG 180
FT
FT Db 5020 AAATCTTCTGCTACAGATGGTACATTGGATGGAATGATGCTTTTGGCTTGTATTCTATG 5079
FT
FT QY 181 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 240
FT
FT Db 5080 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGGAAATA 5139
FT
FT QY 241 CCGGATGATACAGAGGGGCTCCACACTGTTTTAAAGCCGAGATGCGAAACTTTGCG 300
FT


```

Db      5140 CCGATGATACACAGAGGCTCCACACTGTTTAAAGCGGAGATGGCAAACTTTGC 5199
Qy      301 CAAGTTTGCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCTTGC 360
Db      5200 CAAGTTTGCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCTTGC 5259
Qy      361 ATATGCTTATATAGCTGATTCGCACTGAGTGTGTTTACCTGATTTCTCTGTACAGCAA 420
Db      5260 ATATGCTTATATAGCTGATTCGCACTGAGTGTGTTTACCTGATTTCTCTGTACAGCAA 5319
Qy      421 ATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 480
Db      5320 ATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 5379
Qy      481 CTGGCTTCGTTACCAATGCAC 501
Db      5380 CTGGCTTCGTTACCAATGCAC 5400

RESULT 2
ID      AAA89353 standard; DNA; 10329 BP.
AC      AAA89353;
XX
DT      04-NOV-2004 (revised)
DT      23-APR-2001 (first entry)
XX
DE      Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW      Gene silencing; silencing gene; MOM; ds.
XX
OS      Arabidopsis thaliana.
FH      Key
FT      exon      1..1008
FT      /tag= a
FT      /number= 1
FT      /tag= b
FT      /number= 1
FT      1296..2550
FT      /tag= c
FT      /number= 2
FT      1301..9808
FT      /tag= d
FT      /product= "MOM protein"
FT      /note= "this sequence contains 15 introns."
FT      2551..2673
FT      /tag= e
FT      /number= 2
FT      2674..2752
FT      /tag= f
FT      /number= 3
FT      2753..2867
FT      /tag= g
FT      /number= 3
FT      2868..3113
FT      /tag= h
FT      /number= 4
FT      3114..3506
FT      /tag= i
FT      /number= 4
FT      3507..3680
FT      /tag= j
FT      /number= 5
FT      3681..3973
FT      /tag= k
FT      /number= 5
FT      3974..4895
FT      /tag= l
FT      /number= 6
FT      allele
FT      4338

FT      /tag= m
FT      /note= "ecotype Columbia"
FT      4896..4975
FT      /tag= n
FT      /number= 6
FT      4976..5217
FT      /tag= o
FT      /number= 7
FT      5218..5777
FT      /tag= p
FT      /number= 7
FT      5778..5882
FT      /tag= q
FT      /number= 8
FT      5883..6082
FT      /tag= r
FT      /number= 8
FT      6083..7480
FT      /tag= s
FT      /number= 9
FT      6721
FT      /tag= t
FT      /note= "ecotype Columbia"
FT      7481..7615
FT      /tag= u
FT      /number= 9
FT      7616..7771
FT      /tag= v
FT      /number= 10
FT      7772..7914
FT      /tag= w
FT      /number= 10
FT      7915..8070
FT      /tag= x
FT      /number= 11
FT      8071..8153
FT      /tag= y
FT      /number= 11
FT      8154..8318
FT      /tag= z
FT      /number= 12
FT      8319..8451
FT      /tag= aa
FT      /number= 12
FT      8452..8629
FT      /tag= ab
FT      /number= 13
FT      8630..8718
FT      /tag= ac
FT      /number= 13
FT      8719..8918
FT      /tag= ad
FT      /number= 14
FT      8919..9000
FT      /tag= ae
FT      /number= 14
FT      9001..9211
FT      /tag= af
FT      /number= 15
FT      9212..9284
FT      /tag= ag
FT      /number= 15
FT      9285..10329
FT      /tag= ah
FT      /number= 16
FT      /tag= ai
FT      /number= 16
XX      WO200100801-A2.
XX      PD
XX      04-JAN-2001.
XX
XX      21-JUN-2000; 2000WO-EP005761.
XX      PD
XX      23-JUN-1999; 99GB-00014623.
XX      PR

```

XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX Claim 6; Page 19-22; 48pp; English.
XX The present sequence is that of an Arabidopsis thaliana gene (MOM)
CC involved in epigenetic gene silencing. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia, occurring at nucleotide positions
CC 4338 (A instead of T) and 6721 (T instead of G). Gene silencing is useful
CC as a molecular tool for regulating gene expression
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
XX Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 4; Length 10329;
Best Local Similarity 100.0%; Pred. No. 8.9e-133;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATAGTTTAAATTCCTCGATACCTCTAGTGTTCATTTGATTTATGCACTATTTGCTG 60
DB AATAGTTTAAATTCCTCGATACCTCTAGTGTTCATTTGATTTATGCACTATTTGCTG 4959
QY 61 ACTATCTTCCCTACAGGGCTTTGGAGTCCACCAAAAGTCAATACCTCCAGTTCATGT 120
DB ACTATCTTCCCTACAGGGCTTTGGAGTCCACCAAAAGTCAATACCTCCAGTTCATGT 5019
QY 121 AAATCTTCTGGTACAGATGTATCATTTGGATGGAAATGATGCTTTTGGCTTTATTTCTATG 180
DB AAATCTTCTGGTACAGATGTATCATTTGGATGGAAATGATGCTTTTGGCTTTATTTCTATG 5079
QY 181 GGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTTGGGGGAAATYA 240
DB GGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTTGGGGGAAATYA 5139
QY 241 CCGGATGAATCAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGCAAACTTTGC 300
DB CCGGATGAATCAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGCAAACTTTGC 5199
QY 301 CAAATTTTGATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCCTTGC 360
DB CAAATTTTGATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCCTTGC 5259
QY 361 ATATGCTTATATATAGCTGATTTGCACTGATTTGTTACTGATTTCTGTTACAGCCAA 420
DB ATATGCTTATATATAGCTGATTTGCACTGATTTGTTACTGATTTCTGTTACAGCCAA 5319
QY 421 ATGTGAGATTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 480
DB ATGTGAGATTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 5379

QY 481 CTGGCTTCCTTACCAATGCAC 501
DB 5380 CTGGCTTCCTTACCAATGCAC 5400
RESULT 3
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
XX
AC AAA89354;
XX
DT 23-APR-2001 (first entry)
XX Arabidopsis thaliana silencing gene cDNA.
XX Gene silencing; silencing gene; MOM; ss.
XX Arabidopsis thaliana.
XX Key Location/Qualifiers
FT 310..6315
CDS /tag= a
XX WO200100801-A2.
XX 04-JAN-2001.
XX 21-JUN-2000; 2000WO-EP005761.
XX 23-JUN-1999; 99GB-00014623.
XX (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX Claim 6; Page 22-32; 48pp; English.
XX The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression
XX
SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Query Match 48.8%; Score 244.6; DB 4; Length 6571;
Best Local Similarity 98.4%; Pred. No. 2.1e-59;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 69 TCCTACAGGGCTTTGGAGTCCACCAAAAGTCAATACCTCCAGTTCATGTAAATCTTC 128
DB 2973 TCATCTGGGGCTTTGGAGTCCACCAAAAGTCAATACCTCCAGTTCATGTAAATCTTC 3032

QY 129 TGGTACAGATGGTACATGGATGGAATGATGCTTTTGGCTTGATTTCTATGGCAGCCA 188
DB 3033 TGGTACAGATGGTACATGGATGGAATGATGCTTTTGGCTTGATTTCTATGGCAGCCA 3092

QY 189 TATCTCTGGATCCACAGGATATGTTAGCTAGTCAAGATGGGGAAATACCGATGA 248
DB 3093 TATCTCTGGATCCACAGGATATGTTAGCTAGTCAAGATGGGGAAATACCGATGA 3152

QY 249 ATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCAAACTTTGCCAAGTTT 308
DB 3153 ATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCAAACTTTGCCAAGTTT 3212

QY 309 GCATCTTTTCAG 319
DB 3213 GCATCTTTTCAG 3223

RESULT 4
AA89379
ID AAA89379 standard; DNA; 467 BP.
XX
AC AAA89379;
XX
DT 11-SEP-2003 (revised)
DT 23-APR-2001 (first entry)
XX
DE Brassica oleracea silencing gene homologous DNA seq1-43.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Brassica oleracea; acephala.
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
DR WPI; 2001-137952/14.
XX

PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Example 6; Page 45; 48pp; English.
XX

CC This is the sequence of Seq1-43, a genomic DNA sequence of Brassica
CC oleracea acephala that shows homology to the newly identified silencer
CC gene (see AAA89353) of Arabidopsis thaliana. Seq1-27 is a partial
CC sequence of clone 1, which was isolated from a genomic library screened
CC with MOM cDNA. Partial sequences of clone 1 (see also AAA89377-80) showed
CC similarity to different regions of the MOM gene (80-86% at DNA level and
CC 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-
CC terminal parts of the MOM protein (see AAB20062). Gene silencing is
CC useful as a molecular tool for regulating gene expression. (Updated on 11
CC -SEP-2003 to standardise OS field)
XX
SQ Sequence 467 BP; 145 A; 87 C; 108 G; 127 T; 0 U; 0 Other;

Query Match 12.0%; Score 60; DB 4; Length 467;
Best Local Similarity 83.3%; Pred. No. 6.6e-07;
Matches 80; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 AATAGTTTAAATTTCTGCTCCGATACCTCTAGTGTTCATTCATTATGCAAC-TACTTTGCT 59
DB 372 AACAGTTCACTTCTGCTCTCTTACCTCTAGTGTTCATTCATTATGCAATTTTACTTTGCT 431

QY 60 GACTATCTTTCTTACAGGGCTTTGGAGTCACCAAA 95
DB 432 TACTATCTTTCTTACAGGGCATTTGGAGTCACCAAA 467

RESULT 5
AAL41152
ID AAL41152 standard; DNA; 70768 BP.
XX
AC AAL41152;
XX
DT 28-OCT-2002 (first entry)
XX
DE Wooden leg (WOL) gene related DNA SEQ ID No 13.
XX
KW Wooden leg; WOL; vasculature; transgenic plant; agronomic; longer root;
KW wood production; plant; promoter; tree; crop plant; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200244337-A2.
XX
PD 06-JUN-2002.
XX
PF 29-NOV-2001; 2001WO-US045053.
XX
PR 29-NOV-2000; 2000US-0253739P.
XX
PA (UTNY) UNIV NEW YORK STATE.
PA (HELA/) HELARIUTTA Y.
PA (MAHO/) MAHONEN A P.
PA (BONK/) BONKE A W M.
PA (KAUP/) KAUPPINEN L.
PA (RIIK/) RIIKONEN M.
XX
PI Helariutta Y, Mahonen AP, Bonke AMW, Kauppinen L, Riikonen M;
PI Benfey PN;
XX
DR WPI; 2002-599423/64.
XX

PT Novel isolated polypeptide (WOODEN LEG) with ability to regulate a set of
PT asymmetric cell divisions that establish vascular tissue in root and
PT hypocotyl development, useful for improving agronomically valuable
PT plants.
XX
PS Example 1; Page 112-149; 187pp; English.
XX

CC The invention relates to an isolated WOODEN LEG (WOL) polypeptide,
CC comprising 15 contiguous amino acids of a fully defined Arabidopsis
CC WOODEN LEG protein sequence of 1057 amino acids as given in the
CC specification, and to its encoding nucleic acid. The invention also
CC relates to an amino acid sequence of domains of protein, e.g., N-terminal
CC region, C-terminal domain, etc; or is a naturally occurring allelic
CC variant of the above mentioned polypeptide sequence. Expression levels of
CC the nucleic acid can be modified to improve the vasculature in transgenic
CC plants and enhance the agronomic properties of such plants. Also the WOL
CC promoter is used to drive expression of a heterologous coding sequence of
CC trees to improve wood production. The WOL nucleic acid may be used as a
CC molecular marker for a qualitative trait loci, e.g., longer roots or
CC enhanced wood production, in molecular breeding of crop plants. The
CC nucleic acid is also useful in DNA amplification assays to identify the
CC endogenous WOL genes, WOL mutant alleles and/or WOL expression products
CC in cultivars as compared to wild-type plants. They can also be used as
CC markers for linkage analysis of qualitative trait loci. The WOL protein
CC and/or antibodies can be used as diagnostic reagents in immunoassays to
CC detect expression of the WOL gene in cultivars and wild-type plants. The
CC WOL protein, its encoding nucleic acid, and its corresponding antibody
CC are useful for improving agronomically valuable plants e.g., trees. This
CC polynucleotide sequence represents a DNA sequence relating to the wooden
CC leg (WOL) gene of the invention
XX
SQ Sequence 70768 BP; 23616 A; 11681 C; 11519 G; 23952 T; 0 U; 0 Other;

Query Match 8.7%; Score 43.8; DB 6; Length 70768;
Best Local Similarity 51.3%; Pred. No. 0.15;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 281 CGAAGATGCAAACTTGGCCAAAGTTTGCATCTTTCACTAGTGGCCCTTTTTCACCTCC 340
|||||
Db 24845 CGAAATCGCCTTAACAATCGCCATTTTGTGTTTCTGAGAGCTTTTCTTTTAT 24904
|||||

QY 341 ACAACTTATTTAGCTTGCATATGCTATATATAGCTGATTCACACTGTAGTTGTACC 400
|||||
Db 24905 CTAGCTTTTGTATCTCTATTTTATTAAGAATTTTCTATATTAATGTAAT 24964
|||||

QY 401 TGATTTCTCTTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATATCATCCGTTTAA 460
Db 24965 ATTACTATCACCTAGTATATAAAATCTCTAGGAACAATAGCACTCTTTGAA 25024
|||||

QY 461 GCATATTTTATTTCTTATA 479
|||||
Db 25025 ACATTTTATTTTATTTTAA 25043
|||||

RESULT 6
ABN70900/c
ID ABN70900 standard; DNA; 3357 BP.
XX AC ABN70900;
XX
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 9713.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
XX
PN WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR P-PSDB; ABP30269.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 4098; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 3357 BP; 1230 A; 532 C; 660 G; 935 T; 0 U; 0 Other;

Query Match 8.0%; Score 40.2; DB 6; Length 3357;
Best Local Similarity 54.4%; Pred. No. 0.6;
Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 344 ACTTATTTAGCTTGCATATGCTTATATATAGCTGATTCACACTGTAGTTGTACCTGA 403
|||||
Db 2222 ACTTTTCTTGAATTTGAATCTTTATCTTCTTAATTTGAGCAATGGAAGTTGCACTCT 2163
|||||

QY 404 TTTCCTGTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAGCA 463
|||||
Db 2162 TCTTTTAAATAGCAATTTGTTCAAGTTCTTTTCTAAACGATTTTTCATTTTCAGAA 2103
|||||

QY 464 TATTTTATTTCTTATATCTGCTTCGTTA 492
|||||
Db 2102 GTAGTTTGTTCACGCCGCCGCTGCTTAA 2074
|||||

RESULT 7
ABN66624/c
ID ABN66624 standard; DNA; 3537 BP.
XX AC ABN66624;
XX
XX 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide SEQ ID NO 1161.
DE
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR P-PSDB; ABP25993.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3270-3271; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to

CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 3537 BP; 1294 A; 552 C; 707 G; 984 T; 0 U; 0 Other;

Query Match 8.0%; Score 40.2; DB 6; Length 3537;
Best Local Similarity 54.4%; Pred. No. 0.61;
Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 344 ACTTATTTTACGCTTGCATATGCTTATATATAGCTGATGCACTGATGTTGTTACTGA 403
Db 2402 ACTTTTCTTGAATTTGAATCTTTATCTTCTTAATTTGAGCAATTTGAATTTGCAACTCT 2343
Qy 404 TTTCCTGTTACAGCAATGAGAGTTTATCTTCAACTATATCCATCGTTTAAACA 463
Db 2342 TCCTTTTATAGCAAAATGTTCAAGTCTTTTCTTAACGATTTTTCATTTTCAGAA 2283
Qy 464 TATTTTATTTCTTATATCTGCTTCGTTA 492
Db 2282 GTAGTTTGTCCACGCGCGCTGCTTAA 2254

RESULT 8

ABN71527_06/c
Continuation (7 of 22) of ABN71527 from base 600001 (Streptococcus polynucleotide SEQ ID
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

WP	Fragment Name	Begin	End
WP	ABN71527_00	1	110000
WP	ABN71527_01	100001	210000
WP	ABN71527_02	200001	310000
WP	ABN71527_03	300001	410000
WP	ABN71527_04	400001	510000
WP	ABN71527_05	500001	610000
WP	ABN71527_06	600001	710000
WP	ABN71527_07	700001	810000
WP	ABN71527_08	800001	910000
WP	ABN71527_09	900001	1010000
WP	ABN71527_10	1000001	1110000
WP	ABN71527_11	1100001	1210000
WP	ABN71527_12	1200001	1310000
WP	ABN71527_13	1300001	1410000
WP	ABN71527_14	1400001	1510000
WP	ABN71527_15	1500001	1610000
WP	ABN71527_16	1600001	1710000
WP	ABN71527_17	1700001	1810000
WP	ABN71527_18	1800001	1910000
WP	ABN71527_19	1900001	2010000
WP	ABN71527_20	2000001	2110000
WP	ABN71527_21	2100001	2155561

Query Match 8.0%; Score 40.2; DB 6; Length 110000;
Best Local Similarity 54.4%; Pred. No. 1.9;
Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 344 ACTTATTTTACGCTTGCATATGCTTATATATAGCTGATGCACTGATGTTGTTACTGA 403
Db 90383 ACTTTTCTTGAATTTGAATCTTTATCTTCTTAATTTGAGCAATTTGAATTTGCAACTCT 90324
Qy 404 TTTCCTGTTACAGCAATGAGAGTTTATCTTCAACTATATCCATCGTTTAAACA 463
Db 90323 TCCTTTTATAGCAAAATGTTCAAGTCTTTTCTTAACGATTTTTCATTTTCAGAA 90264
Qy 464 TATTTTATTTCTTATATCTGCTTCGTTA 492

Db 90263 GTAGTTTGTCCACGCGCGCTGCTTAA 90235

RESULT 9

ABK40092
ID ABK40092 standard; DNA; 9728 BP.

XX AC ABK40092;

XX 21-MAY-2002 (first entry)

XX Human chemically pretreated gene sequence #87 strand 2.

XX Human; db; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
XX Cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
XX UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

XX Homo sapiens.

XX WO200202806-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP007470.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-154757/20.

XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligoners,
XX useful for detecting cytosine methylation state of genes associated with
XX pharmacogenomics and for therapy of diseases e.g. cancer.

Claim 1; SEQ ID NO 174; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
CC NM_019900, NM_019901, NM_019902, NM_019863, NM_019898, NM_019899, and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 U; 0 Other;

Query Match 7.8%; Score 39.2; DB 6; Length 9728;
Best Local Similarity 56.1%; Pred. No. 1.6;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 349 TTTTAGCCTTGCATATGCTTATATATAGCTGATGCAACTGTAGTTGTTACCTGATTTC 408

```
Db      701 TTTTAAATTAGGAAGTTTATATAATATAGATTAATGTGCAATTTGTTATTTTATAAATTTT 760
QY      409 TGTACAGCAATGAGAGTTTATCTTCAACTATATATCCATCCGTTTAAAGCATATTT 468
Db      761 TTTTATATAAATTAATTAATTTAAATGTTAGGATATATATTTTATTAAAGATTTT 820
QY      469 TATTTCTTATAT 480
Db      821 TATTTTTTTTTT 832

RESULT 10
ABL33903
ID ABL33903 standard; DNA; 9728 BP.
XX
XX
AC ABL33903;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1876.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemia; cytosine methylation; antiasthmatic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
XX Claim 1; SEQ ID NO 1876; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
XX Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 U; 0 Other;
XX
XX Query Match 7.8%; Score 39.2; DB 6; Length 9728;
XX Best Local Similarity 56.1%; Pred. No. 1.6;
XX Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY      349 TTTTACCTTGCATATGCTTATATAGCTGATTGCAACTGTTGTTACCTGATTTC 408
Db      701 TTTTAAATTAGGAAGTTTATATAATATAGATTAATGTGCAATTTGTTATTTTATAAATTTT 760
```

```
QY      409 TGTACAGCAATGAGAGTTTATCTTCAACTATATATCCATCCGTTTAAAGCATATTT 468
Db      761 TTTTATATAAATTAATTAATTTAAATGTTAGGATATATATTTTATTAAAGATTTT 820
QY      469 TATTTCTTATAT 480
Db      821 TATTTTTTTTTT 832

RESULT 11
ABL34608
ID ABL34608 standard; DNA; 18585 BP.
XX
XX ABL34608;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human metastasis associated gene SEQ ID NO: 161.
XX
XX Metastasis associated gene; cytosine methylation; gene therapy; cancer;
KW cytosine methylation; gene; ds.
XX
XX Homo sapiens.
XX
XX WO200177376-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP003970.
XX
XX 06-APR-2000; 2000DE-01019058.
PR
XX 07-APR-2000; 2000DE-01019173.
PR
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-010922/01.
XX
XX New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
XX Claim 1; SEQ ID NO 161; 23pp + Sequence Listing; English.
XX
XX The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention
XX
XX Sequence 18585 BP; 5934 A; 120 C; 3607 G; 8924 T; 0 U; 0 Other;
XX
XX Query Match 7.6%; Score 38.2; DB 6; Length 18585;
XX Best Local Similarity 48.0%; Pred. No. 4;
XX Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY      272 TTTTAAAGCGAAGATGGCAAAACTTTGCAAGTTTTCATCTTTTCAGTAAGTGGCTTT 331
Db      8988 TTTTATAGTAAGGTGATTAATTAATTTAAATATAGGTAGTAGGAGGAATTT 9047
QY      332 TTCACCTCCACAACTTATTTTAGCCTTGCATATGCTTATATATAGTGAATTCGAATGTA 391
Db      9048 TTGAGGTTAGGAAGGTTTTTATATTTTTCGTTATGTTAGTTTTCGAGTTTTTA 9107
QY      392 GTTGTACCTGATTTCTCTTACAGCAATGTGAGAGTTTATTTCTTCACTATATATCCA 451
Db      9108 GTTGGAAATTTTAAATTTTATAGGAATTAATTTGTTATATATTTTAAATGTTTAT 9167
QY      452 TCGTTTAAAGCATATTTTATTTTCTTATATCTGCTTCGTTACCAATG 498
Db      9168 TAAGTAAATTTTAATGATTTTATGATGTTGTTTATAGATG 9214
```


PS Claim 1; SEQ ID NO 62; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6283 BP; 2148 A; 106 C; 1340 G; 2689 T; 0 U; 0 Other;
Query Match 7.6%; Score 38; DB 6; Length 6283;
Best Local Similarity 50.0%; Pred. No. 3.1;
Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 291 AAACCTTCCCAAGTTTGCATCTTCAGTAAGTGGCCTTTTCACCTCCACACTATT 350
DB 3735 AATCATTTAAAATAATACCATTAACGATATAAATCTTTCAATCCAAAACATAAA 3676
QY 351 TTAGCCTTCATATGCTTATATAGCTGATTCGAATGAGTGTGTACTGATTTCCCTG 410
DB 3675 ATATCAATTCATTTATTAATCACTATTATTCACAAATCTTTTATTATTATTCAAA 3616
QY 411 TTACAGCCAAATGTGAGAGTTTATTCTTCAACTATATCCATCGTTTAAGCATATTTTA 470
DB 3615 CTACAAATTTAAATCCTTTTATTAAATTTATTTCTTAAATTTTATTATTATTATTTA 3556
QY 471 TTTCTTATAT 480
DB 3555 TTTTATTAAT 3546
RESULT 14
ADC99111/c
ID ADC99111 standard; cDNA; 3987 BP.
XX
AC ADC99111;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human KPP cDNA - SEQ ID 64.
XX
KW anti-HIV; antiallergic; antiinflammatory; antianaemic; antiparkinsonian;
KW neurotic; anticonvulsant; antiarteriosclerotic; antiasthmatic;
KW immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;
KW antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;
KW osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoiatric;
KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;
KW viricide; protozoicide; fungicide; kinase; phosphatase; KPP;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis;
KW cancer; developmental; mental retardation; neurological;
KW Alzheimer's disease; Parkinson's; autoimmune; inflammatory; Crohn's;
KW diabetes mellitus; viral; bacterial; fungal; parasitic; protozoan;
KW helminthic infection; transgenic; gene therapy; human; ss; gene.
XX
OS Homo sapiens.
XX
PN WO2003033680-A2.
XX
PD 24-APR-2003.
XX
PF 17-OCT-2002; 2002WO-US033723.
XX
PR 19-OCT-2001; 2001US-0345474P.
PR 02-NOV-2001; 2001US-0343910P.
PR 13-NOV-2001; 2001US-0333098P.
PR 16-NOV-2001; 2001US-0332424P.
PR 30-NOV-2001; 2001US-0334288P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX

PI Bandman O, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
PI Emerling EM, Forsythe IJ, Gandhi AR, Gorvad AE, Griffin JA;
PI Gururajan R, Hafalia AJA, Khan FA, Lal PG, Lee EA, Lee SY;
PI Lindquist EA, Lu DAM, Lu Y, Marquis JP, Nguyen DB, Arvizu CS;
PI Ramkumar J, Recipon SA, Richardson TW, Swarnakar A, Tang YT;
PI Thornton MB, Tran UK, Chawla NK, Warren BA, Yang J, Yao MG, Yue H;
PI Zebarjadian Y;
XX
DR WPI; 2003-403214/38.
DR P-PSDB; ADC99059.
XX
PT New human kinases and phosphatases and polynucleotides, useful for
PT diagnosing, treating or preventing autoimmune or inflammatory disorders
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
PT cancer or hepatitis.
XX
PS Claim 5; SEQ ID NO 64; 424pp; English.
XX
CC The invention relates to a novel isolated polypeptide which is a human
CC kinase and phosphatase (KPP). The KPP polypeptides, polynucleotides,
CC agonists and antagonists are useful for diagnosing, treating or
CC preventing cell proliferative disorders such as atherosclerosis,
CC cirrhosis, hepatitis and cancer, developmental disorders e.g. mental
CC retardation, neurological disorders including Alzheimer's disease and
CC Parkinson's disease, autoimmune and inflammatory disorders such as
CC Crohn's disease and diabetes mellitus and finally, viral, bacterial,
CC fungal, parasitic, protozoan or helminthic infections. Furthermore, the
CC polynucleotides encoding KPP may be useful for creating transgenic
CC animals to model human disease, as well as during gene therapy
CC procedures. The current sequence is that of the human KPP cDNA of the
CC invention.
XX
SQ Sequence 3987 BP; 1347 A; 760 C; 880 G; 1000 T; 0 U; 0 Other;
Query Match 7.5%; Score 37.4; DB 10; Length 3987;
Best Local Similarity 52.2%; Pred. No. 4;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 231 GGGGAAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGC 290
DB 3627 GGAATAATACAAACCCAACTTAGATGAAGGCCAAAATTTAAATATCGAAAACATC 3568
QY 291 AAAACCTTCCCAAGTTTTCATCTTCAGTAAGTGGCCTTTTCACCTCCACACTATT 350
DB 3567 GAAACTGTGCACATCAAAAATGTTGAGTAACCTGTTCTGTGAGAGCTGAAAGTTTT 3508
QY 351 TTAGCCTTCATATGCTTATATATATAGCTGATTGCAACTG 389
DB 3507 GTTGACACAAAAGTATTATGTACAACTAAGCTTACTG 3469
RESULT 15
ADH29035/c
ID ADH29035 standard; DNA; 4613 BP.
XX
AC ADH29035;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human chronic myelogenous leukaemia (CML) gene marker #303.
XX
KW ds; chronic phase chronic myelogenous leukaemia; CP-CML;
KW blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
KW gene marker.
XX
OS Homo sapiens.
XX
PN US2003104426-A1.
XX
PD 05-JUN-2003.
XX
PF 14-JUN-2002; 2002US-00171581.
XX


```
PR 18-JUN-2001; 2001US-0298914P.
XX
PA (LINS/) LINSLEY P S.
PA (MAOM/) MAO M.
PA (DAIH/) DAI H.
PA (HEYV/) HE Y.
PA (RADI/) RADICH J P.
XX
PI Linsley PS, Mao M, Dai H, He Y, Radich JP;
XX
XX WPI; 2003-787046/74.
XX
XX Classifying cell sample as chronic phase myelogenous leukemia or
PT blast crisis chronic myelogenous leukemia by detecting difference in
PT expression of genes corresponding to the markers such as X15415, U89436.
XX
XX PS Disclosure; SEQ ID NO 303; 31pp; English.
XX
XX The invention relates to a method of classifying a cell sample as chronic
CC phase chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-
CC CML). The method is useful for classifying a sample as CP-CML or BC-CML.
CC The present sequence represents a human chronic myclogenous leukaemia
CC (CML) gene marker used to distinguish blast crisis CML from chronic phase
CC CML.
XX
SQ Sequence 4613 BP; 1579 A; 831 C; 965 G; 1238 T; 0 U; 0 Other;
Query Match 7.5%; Score 37.4; DB 10; Length 4613;
Best Local Similarity 52.2%; Pred. No. 4.2;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 231 GGGGAAATACCGGATGAATCAGAGAGGCGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 3589 GGAATAATACAAACCCCAACTTAGATGAAAGCCAAATAATTTAAATATCGAAAACTC 3530
QY 291 ABACTTTGCCAGTTTTCATCTTCAGTAAGTGGCTTTTCACCTCCACAACTATT 350
Db 3529 GAAACTGTGGCAGCATCAAAAATGTTGAGTAACTGTTCTCTGTGAGAGCTGAAAAGTTTT 3470
QY 351 TTAGCCTTGCTATGCTTATATATAGCTGATTGCAACTG 389
Db 3469 GTTGACACAAAGTATTATGTAACACTAGGCTTACTG 3431
```

Search completed: July 12, 2005, 01:45:54
Job time : 388.755 secs

THIS PAGE BLANK (USPTO)

Db 458 RAWDHKCTHNNNTTWKMTYNNCYWKNWNGKSHRBAAAVVTTWMMWRRYAHANNNN 517
QY 363 ATGCTTATATATAGCTGATTGCAACTGTAGTCTTACCTGATTCTCTGTACAGCCAAAT 422
Db 518 WDWKACATWYKVBVCWKNNYAAWYTKSSWNTSRVYRWKNNNSWRWSRSDTRSGRAN 577
QY 423 GTGAGAGTTTATCTTCAACATATATCCATCCGTTTAAAGCATATTTTATTTCTT 476
Db 578 NVARABHYGKWNTRWBSHTWBHBRAGAAYHMMBMYBAKCHCMKAWYKAKK 631

RESULT 2

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F18
US-08-232-463-14

Query Match 7.4%; Score 37.2; DB 1; Length 7218;
Best Local Similarity 10.4%; Pred. No. 1;
Matches 24; Conservative 114; Mismatches 92; Indels 0; Gaps 0;
QY 265 CACACTGTTTAAAGCGGAGATGGCAAACTTTGCCAGTTTTCGATCTTTTCAGTAAGT 324
Db 1018 CTCACAGAATAATTCGAGCTTGGCTGCGAGGTCGAGGGAGCTTGCATATTTTTTTTT 1077
QY 325 GGCCTTTTTCACCTCCCACTTATTTTACCTTTCGATGCTTATATATAGCTGATTCG 384
Db 1078 YY 1137
QY 385 AACTGTAGTTGTACTGATTTCTCTGTACAGCCAAATGTGAGAGTTTTATTTCTCACT 444

Db 1138 YY 1197
QY 445 ATATCATCCGTTTAAAGCATATTTTATTTCTTATATCTCGCTTCGTTACC 494
Db 1198 YY 1247
RESULT 3
US-09-453-702B-72
; Sequence 72, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27396
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46819
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-453-702B-72

Query Match 7.4%; Score 37.2; DB 3; Length 46819;
Best Local Similarity 49.5%; Pred. No. 2.1;
Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 298 TGGCAAGTTTTCGATCTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTTATTTAGCCT 357
Db 36780 TGGCAGGCTGTAATTTGACATCATGAGTAGCTTTTCTCTTCCAGTGTCTTTTCAGGAG 36839
QY 358 TGCATATGCTTATATATAGCTGATTCGAACTGTAGTTGTTACCTGATTTTCTGTTACAGC 417
Db 36840 TTTAGCCGCTTTTATATATCTTAATGTCTCAGGAGTGTTCACAAAGTGCATACTGGC 36899
QY 418 CAAATGTGAGAGTTTATTTCTTCAACTATATCATCCGTTTAAAGCATATTTTATTTCTTA 477
Db 36900 AATTTTCATAGTGTCTTACTTTTGTGCTGCAGCAATCAATGCAGCATCAGTACTTCATT 36959
QY 478 TATCTGGCTTCGTT 491

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 543.528 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929C-1_COPY_4900_5400

Perfect score: 501

Sequence: 1 aatagtttaattctgtcc.....tggcttcgtaccatgcac 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 23: /cgn2_6/prodata/2/pubpna/US11A_PUBCOMB.seq:*
- 24: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.8	8.7	70768	US-10-135-322-13	Sequence 13, Appl
2	39.2	7.8	9728	US-10-311-455-1876	Sequence 1876, Ap
3	39.2	7.8	9728	US-10-257-166-174	Sequence 174, App
4	38.2	7.6	18585	US-10-240-485-161	Sequence 161, App
5	38	7.6	201	US-10-719-993-38032	Sequence 38032, A
6	38	7.6	6283	US-10-311-455-62	Sequence 62, Appl
7	37.6	7.5	86474	US-10-719-993-6879	Sequence 6879, Ap

c 8	37.4	7.5	4613	15	US-10-171-581-303	Sequence 303, App
c 9	37.4	7.5	4613	21	US-10-887-553A-343	Sequence 343, App
c 10	37.4	7.5	5704	19	US-10-741-601-204	Sequence 204, App
c 11	37.4	7.5	6016	19	US-10-741-601-203	Sequence 203, App
c 12	37.4	7.5	6196	19	US-10-741-601-202	Sequence 202, App
c 13	37.4	7.5	179041	19	US-10-741-601-5678	Sequence 5678, Ap
c 14	37.2	7.4	46819	14	US-10-114-170-72	Sequence 72, Appl
c 15	37.2	7.4	367378	16	US-10-312-841-1	Sequence 1, Appl
c 16	37	7.4	671	10	US-09-814-353-246	Sequence 246, App
c 17	37	7.4	671	10	US-09-814-353-6636	Sequence 6636, Ap
c 18	37	7.4	367378	16	US-10-312-841-2	Sequence 2, Appl
c 19	36.8	7.3	701	15	US-10-204-085-15	Sequence 15, Appl
c 20	36.8	7.3	769	13	US-10-027-632-16766	Sequence 16766, A
c 21	36.8	7.3	769	13	US-10-027-632-16767	Sequence 16767, A
c 22	36.8	7.3	769	13	US-10-027-632-146401	Sequence 146401, A
c 23	36.8	7.3	769	13	US-10-027-632-146402	Sequence 146402, A
c 24	36.8	7.3	769	17	US-10-027-632-16766	Sequence 16766, A
c 25	36.8	7.3	769	17	US-10-027-632-16767	Sequence 16767, A
c 26	36.8	7.3	769	17	US-10-027-632-146401	Sequence 146401, A
c 27	36.8	7.3	769	17	US-10-027-632-146402	Sequence 146402, A
c 28	36.8	7.3	1204	15	US-10-204-085-13	Sequence 13, Appl
c 29	36.8	7.3	15620	9	US-09-928-457-36	Sequence 36, Appl
c 30	36.4	7.3	201	20	US-10-719-993-38034	Sequence 38034, A
c 31	36.4	7.3	650	13	US-10-027-632-67708	Sequence 67708, A
c 32	36.4	7.3	650	13	US-10-027-632-311474	Sequence 311474, A
c 33	36.4	7.3	650	17	US-10-027-632-67708	Sequence 67708, A
c 34	36.4	7.3	650	17	US-10-027-632-311474	Sequence 311474, A
c 35	36.2	7.2	127369	13	US-10-087-192-238	Sequence 238, App
c 36	36	7.2	2526	17	US-10-374-780A-1153	Sequence 1153, App
c 37	36	7.2	2526	18	US-10-412-6998B-1352	Sequence 1352, Ap
c 38	35.8	7.1	614	18	US-10-424-599-95239	Sequence 95239, A
c 39	35.8	7.1	96596	11	US-09-997-722-196	Sequence 196, App
c 40	35.8	7.1	175590	10	US-09-911-077A-13	Sequence 13, Appl
c 41	35.8	7.1	175590	20	US-10-724-806-13	Sequence 13, Appl
c 42	35.6	7.1	1954	20	US-10-425-115-117144	Sequence 117144, A
c 43	35.4	7.1	49600	18	US-10-459-262A-2	Sequence 2, Appl
c 44	35.4	7.1	130349	21	US-10-741-600-17619	Sequence 17619, A
c 45	35.4	7.1	138363	19	US-10-367-094-117	Sequence 117, App

ALIGNMENTS

RESULT 1
US-10-135-322-13
; Sequence 13, Application US/10135322
; Publication No. US20020173017A1
; GENERAL INFORMATION:
; APPLICANT: BENFEY, PN
; APPLICANT: HELARIUTTA, Y
; APPLICANT: MAHONEN, AP
; APPLICANT: BONKE, AWM
; APPLICANT: KAUPPINEN, L
; APPLICANT: RIIKONEN, M
; TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF
; FILE REFERENCE: 5914-086-999
; CURRENT APPLICATION NUMBER: US/10/135,322
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: 60/253,739
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 70768
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-135-322-13

Query Match 8.7%; Score 43.8; DB 13; Length 70768;
Best Local Similarity 51.3%; Pred. No. 0.48;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
Qy 281 CGAAGATGCGAANAACCTTGCCAAAGTTTGCATCTTTCAGTAAGTGGCCTTTTTCACCTCC 340

Db 24845 CGAAATCGCCTTAACAATCGCCATTTTGTGTTTCTGAGAGCTTTTTTTTTTAT 24904
QY 341 ACAACTATTATTTAGCCTTGATATGCTATATAGCTGATTCGAACCTGTAGTTGTACC 400
Db 24905 CTAGCTTTTGTATCTCTATTTTATTAAGAATTTTCTATATTAATGTAAT 24964
QY 401 TGATTTCTGTGTACGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCCGTTTAA 460
Db 24965 ATTACTATCACCTAGTATAAATAAATCTCTAGGAACAATAGCACTCTTTGAA 25024
QY 461 GCATATTTATTTCTTATA 479
Db 25025 ACATTTTATTTTATTTT 25043

RESULT 2
US-10-311-455-1876
; Sequence 1876, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1876
; LENGTH: 9728
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1876

Query Match 7.8%; Score 39.2; DB 15; Length 9728;
Best Local Similarity 56.1%; Pred. No. 3.9;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 349 TTTTAGCCTTGATATGCTTATATAGCTGATTCGAACCTGTAGTTGTACCTGATTTCC 408
Db 701 TTTTAATTAGGAAGTTTATAAATATAGATTATTGTGTAATTTGTTATTTATTAATTTT 760
QY 409 TGTACAGCAATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTT 468
Db 761 TTTTATATAAATATTTATTTAATGTTAGGATATATATTTTATTTAAGATTATTT 820
QY 469 TATTTCTTATAT 480
Db 821 TATTTTATTTT 832

RESULT 3
US-10-257-166-174
; Sequence 174, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166

; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 174
; LENGTH: 9728
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-174

Query Match 7.8%; Score 39.2; DB 17; Length 9728;
Best Local Similarity 56.1%; Pred. No. 3.9;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 349 TTTTAGCCTTGATATGCTTATATAGCTGATTCGAACCTGTAGTTGTACCTGATTTCC 408
Db 701 TTTTAATTAGGAAGTTTATAAATATAGATTATTGTGTAATTTGTTATTTATTAATTTT 760
QY 409 TGTACAGCAATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTT 468
Db 761 TTTTATATAAATATTTATTTAATGTTAGGATATATATTTTATTTAAGATTATTT 820
QY 469 TATTTCTTATAT 480
Db 821 TATTTTATTTT 832

RESULT 4
US-10-240-485-161
; Sequence 161, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 161
; LENGTH: 18585
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-161

Query Match 7.6%; Score 38.2; DB 15; Length 18585;
Best Local Similarity 48.0%; Pred. No. 10;
Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 272 TTTTAAAGCCGAAGATGGCAAACTTTTGCAGATTTTGCATCTTTTCAGTAGTGGCTTT 331
Db 8988 TTTTATGTTAAGAGGTGATTAATATTTTAAATTTAAATATAGGTAGTAGGAGGAATTT 9047

QY 332 TTCACTCCCAACTATTTAGCTTGCATATGCTTATATATAGCTGATGCAACTGTA 391
Db 9048 TTGAGGTGAGAAAGTTTATATTTATTTTTCGTTTATGTTAGTTGATTTT 9107
QY 392 GTTGTACCTGATTTCTCTGTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCA 451
Db 9108 GTTGGAAATTTATAATTTTATAGGAAATAGTAATTTGTTATATTTTAAGTTTAT 9167
QY 452 TCGTTTAAACATATTTATTTCTTATATCTGCTTGGCTTACCAATG 498
Db 9168 TAAGTAAAAATTTAATGATTTTATGATGCTGGTTTTTATATAGATG 9214

RESULT 5
US-10-719-993-38032/c
; Sequence 38032, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38032
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-38032

Query Match 7.6%; Score 38; DB 20; Length 201;
Best Local Similarity 59.6%; Pred. No. 1.3;
Matches 62; Conservative 1; Mismatches 41; Indels 0; Gaps 0;
QY 349 TTTTAGCTTGCATATGCTTATATATAGCTGATGCAACTGTAGTTGTTACCTGATTTCC 408
Db 146 TTATATATTATATACATAAATATAAATGCTTACAAATAATATKTTATATATGATTTCC 87
QY 409 TGTACAGCAATGTGAGAGTTTATTTCTTCAACTATATCCAT 452
Db 86 TAATATAAACAAAAATCAATAATCCTGTTCTTCTACTGTCTAT 43

RESULT 6
US-10-311-455-62/c
; Sequence 62, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 62
; LENGTH: 6283
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-62

Query Match 7.6%; Score 38; DB 15; Length 6283;
Best Local Similarity 50.0%; Pred. No. 7;
Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 291 AAAAAGTTTGCAGTTTTCATCTTTTCAGTAAGTGGCTTTTTCACCTCCCAACTATT 350
Db 3735 AATCATTTAAAAATTAATACCATTAATACGATTAATTAATCTTCAATCCAAACATAAA 3676
QY 351 TTAGCTTGCATATGCTTATATATAGCTGATGCAACTGTAGTTGTTACCTGATTTCC 410
Db 3675 ATATCATTCCTATTTATTTAAATCTAGCTTATTTCAACAATTTCTTTTATTTTCAAA 3616
QY 411 TTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCATCGTTTACGATATTTTA 470
Db 3615 CTACAAATTTTAAATCCTTTTATTAATTTTCTTAAATTTTATTTTATTTTATTTTA 3556
QY 471 TTTCTTATAT 480
Db 3555 TTTTATTAAT 3546

RESULT 7
US-10-719-993-6879/c
; Sequence 6879, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6879
; LENGTH: 86474
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-6879

Query Match 7.5%; Score 37.6; DB 20; Length 86474;
Best Local Similarity 58.7%; Pred. No. 34;
Matches 61; Conservative 2; Mismatches 41; Indels 0; Gaps 0;
QY 349 TTTTAGCTTGCATATGCTTATATATAGCTGATGCAACTGTAGTTGTTACCTGATTTCC 408
Db 41296 TTATATATTATATACATAAATATAAATGCTTACAAATAATATKTTATATATGATTTCC 41237
QY 409 TGTACAGCAATGTGAGAGTTTATTTCTTCAACTATATCCAT 452
Db 41236 TAATATAAACAAAAATCAATAATCCTGTTTTCATCTACTGTCTAT 41193

RESULT 8
US-10-171-581-303/c
; Sequence 303, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 303
; LENGTH: 4613
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:

```
; DATABASE ACCESSION NUMBER: D87930
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-303

Query Match
Best Local Similarity 7.5%; Score 37.4; DB 15; Length 4613;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 3589 GGAAAAAATACAAACCAACTTGTAGATGAAAGCCAAAAAATTTAAATATCGAAAAAATC 3530

QY 291 AAAAATTTGCCAAGTTTTCATCTTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTTATT 350
Db 3529 GAAACTGTGGCAGCATCAAAAAATGTTGAGTAACCTGCTCTCTGTGAGAGCTGAAAGTTTTT 3470

QY 351 TTAGCCTTCATATGCTTATATATAGCTGATTGCAACTG 389
Db 3469 GTTGACACAAAAGTATTATGTACAACTAAGGCTTACTG 3431

RESULT 9
US-10-887-553A-343/c
; Sequence 343, Application US/10887553A
; Publication No. US20050085436A1
; GENERAL INFORMATION:
; APPLICANT: Li, Hao
; TITLE OF INVENTION: Method to treat conditions associated
; with insulin signalling dysregulation
; FILE REFERENCE: 4-33262
; CURRENT APPLICATION NUMBER: US/10/887,553A
; PRIOR FILING DATE: 2004-07-08
; PRIOR APPLICATION NUMBER: 60/485,883
; NUMBER OF SEQ ID NOS: 1208
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 343
; LENGTH: 4613
; TYPE: DNA
; ORGANISM: human
US-10-887-553A-343

Query Match
Best Local Similarity 7.5%; Score 37.4; DB 21; Length 4613;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 3589 GGAAAAAATACAAACCAACTTGTAGATGAAAGCCAAAAAATTTAAATATCGAAAAAATC 3530

QY 291 AAAAATTTGCCAAGTTTTCATCTTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTTATT 350
Db 3529 GAAACTGTGGCAGCATCAAAAAATGTTGAGTAACCTGCTCTCTGTGAGAGCTGAAAGTTTTT 3470

QY 351 TTAGCCTTCATATGCTTATATATAGCTGATTGCAACTG 389
Db 3469 GTTGACACAAAAGTATTATGTACAACTAAGGCTTACTG 3431

RESULT 10
US-10-741-601-204/c
; Sequence 204, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 28415
; SOFTWARE: FastSeq for Windows Version 4.0
US-10-741-601-204/c

; SEQ ID NO 204
; LENGTH: 5704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-204

Query Match
Best Local Similarity 7.5%; Score 37.4; DB 19; Length 5704;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 3841 GGAAAAAATACAAACCAACTTGTAGATGAAAGCCAAAAAATTTAAATATCGAAAAAATC 3782

QY 291 AAAAATTTGCCAAGTTTTCATCTTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTTATT 350
Db 3781 GAAACTGTGGCAGCATCAAAAAATGTTGAGTAACCTGCTCTCTGTGAGAGCTGAAAGTTTTT 3722

QY 351 TTAGCCTTCATATGCTTATATATAGCTGATTGCAACTG 389
Db 3721 GTTGACACAAAAGTATTATGTACAACTAAGGCTTACTG 3683

RESULT 11
US-10-741-601-203/c
; Sequence 203, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203
; LENGTH: 6016
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-601-203

Query Match
Best Local Similarity 7.5%; Score 37.4; DB 19; Length 6016;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 4331 GGAAAAAATACAAACCAACTTGTAGATGAAAGCCAAAAAATTTAAATATCGAAAAAATC 4272

QY 291 AAAAATTTGCCAAGTTTTCATCTTTTCAGTAAGTGGCCTTTTTCACCTCCACAACTTATT 350
Db 4271 GAAACTGTGGCAGCATCAAAAAATGTTGAGTAACCTGCTCTCTGTGAGAGCTGAAAGTTTTT 4212

QY 351 TTAGCCTTCATATGCTTATATATAGCTGATTGCAACTG 389
Db 4211 GTTGACACAAAAGTATTATGTACAACTAAGGCTTACTG 4173

RESULT 12
US-10-741-601-202/c
; Sequence 202, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
```

```
; ; LENGTH: 6196
; ; TYPE: DNA
; ; ORGANISM: Homo sapiens
US-10-741-601-202

Query Match          7.5%; Score 37.4; DB 19; Length 6196;
Best Local Similarity 52.2%; Pred. No. 10; Mismatches 0; Gaps 0;
Matches 83; Conservative 0; Indels 76;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 4333 GGAATAATACAAACCCAACTTAGATGAAGGCAAAATTTAAATATCGAAAACTC 4274

QY 291 AAAACTTGGCAAGTTTGGCATCTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATT 350
Db 4273 GAAACTGTGGCAGCATCAAAAATGTTGAGTAACCTGCTCTGTGAGAGCTGAAAGTTT 4214

QY 351 TTAGCCTTGCATATGCTTATATATAGCTGATTGCAACTG 389
Db 4213 GTTGACAAAAGTATTATGTACAACCTAAGGCTTACTG 4175

RESULT 13
US-10-741-601-5678/c
; Sequence 5678, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 5678
; LENGTH: 179041
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(179041)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5678

Query Match          7.5%; Score 37.4; DB 19; Length 179041;
Best Local Similarity 52.2%; Pred. No. 55;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 231 GGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTAAAGCCGAAGATGGC 290
Db 171179 GGAATAATACAAACCCAACTTAGATGAAGGCAAAATTTAAATATCGAAAACTC 171120

QY 291 AAAACTTGGCAAGTTTGGCATCTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATT 350
Db 171119 GAAACTGTGGCAGCATCAAAAATGTTGAGTAACCTGCTCTGTGAGAGCTGAAAGTTT 171060

QY 351 TTAGCCTTGCATATGCTTATATATAGCTGATTGCAACTG 389
Db 171059 GTTGACAAAAGTATTATGTACAACCTAAGGCTTACTG 171021

RESULT 14
US-10-114-170-72
; Sequence 72, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Perna, Valerie
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
```

```
; ; NUMBER OF SEQUENCES: 265
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Quarles & Brady
; ; STREET: 1 South Pinckney Street
; ; CITY: Madison
; ; STATE: WI
; ; COUNTRY: US
; ; ZIP: 53701-2113
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: Word Perfect 8.0
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/10/114,170
; ; FILING DATE: 01-Apr-2002
; ; CLASSIFICATION: <Unknown>
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: 09/453,702
; ; FILING DATE: 03-DEC-1999
; ; APPLICATION NUMBER: 60/110,955
; ; FILING DATE: 04-DEC-1998
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Seay, Nicholas J.
; ; REGISTRATION NUMBER: 27386
; ; REFERENCE/DOCKET NUMBER: 960296.95017
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (608) 251-5000
; ; TELEFAX: (608) 251-9166
; ; INFORMATION FOR SEQ ID NO: 72:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 46819
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: DNA (genomic)
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-10-114-170-72

Query Match          7.4%; Score 37.2; DB 14; Length 46819;
Best Local Similarity 49.5%; Pred. No. 32;
Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 298 TGGCAAGTTTGGCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCCT 357
Db 36780 TGGCAGGTCTGTTAATTGACATCAGTAGCTTTTCTCTTCCAGTGTTTCTTTTCAGGAG 36839

QY 358 TGCATATGCTTATATATAGCTGATTCGAACTGATGTTGTTACCTGATTTCTGTTACAGC 417
Db 36840 TTTAGCCGCTTTTATATATCTTAATTGCTCAGCGAGTGTACCAAAAGTGCCATAACTGGC 36899

QY 418 CAAATGTGAGAGTTTTATTTCTTCAACTATATATCATCCGTTTAAAGCATATTTTATTTCTTA 477
Db 36900 AATTTCATAGTCTCTACTCTTTTGTGCTGAGCAATCAATGCAATCATCAGTACTTCATT 36959

QY 478 TATCTGGCTTCGTT 491
Db 36960 TTTATCAGTGCTTT 36973

RESULT 15
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: EpiGenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MH
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match      7.4%; Score 37.2; DB 16; Length 3673778;
Best Local Similarity 53.4%; Pred. No. 2.6e+02;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 346 TTATTTAGCCTTGCATATGCTTATATATAGCTGATTGCAACTGTAGTTGTTACCTGATT 405
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1097433 TAAGTTTGTATTATTATTATTATTTTGTGTTTGTAGTGGTTGAATTTTAAAT 1097492
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 406 TCCTGTTACAGCCAAATGTGAGAGTTTATCTTCAACTATATCCATCCGTTTAAGCATA 465
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1097493 TTTCGTTATAGATAAAGGTTTAGGTTTTTTTTTTTAAATTTTATCGTGTATTTT 1097552
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 466 TTTTATTTCTTATATCTGGCTTCGTT 491
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1097553 TTTTATTTTATATATATTTTCGTT 1097578
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 12, 2005, 11:12:17
Job time : 552.528 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005: Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:03:44 ; Search time 2468.64 Seconds
(without alignments)
7724.974 Million cell updates/sec

Title: US-10-018-929C-1_COPY_4900_5400

Perfect score: 501
Sequence: 1 aatagtttaattctgtctcc.....tggcttcgttaaccatgcac 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479089

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hsc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	54.4	10.9	342	7	CV262174 WS02018.B
2	43.2	8.6	732	9	AG415172 Mus muscu
3	42.8	8.5	1279	9	AG310706 Mus muscu
4	42.2	8.4	1101	9	AL063921 Drosophil
C 5	41.8	8.3	714	5	BUS70551 AGENCOURT
C 6	41.4	8.3	391	2	BE045401 hn21f08.x
7	41.2	8.2	449	7	CK520898 rswea0.00
C 8	41.2	8.2	587	6	CD681805 rj31d01.y
C 9	41.2	8.2	638	6	CD682973 rj28g12.y
C 10	40.6	8.1	824	7	CK310894 S802007B1
11	40.4	8.1	1316	9	AG390927 Mus muscu
12	40	8.0	816	3	AF173385 Homo sapi
13	39.8	7.9	1188	9	AG305042 Mus muscu
14	39.8	7.9	1476	9	CL647966 CH213-157
15	39.4	7.9	774	5	BUS37425 603473713
16	39.4	7.9	1101	9	CNS00FNX
17	39.4	7.9	1134	9	AG305361 Mus muscu
18	39.2	7.8	704	8	BZ248685 CH230-363
19	39.2	7.8	2270	9	AG279272 Mus muscu
20	39	7.8	322	9	CNS01W9Y
21	39	7.8	731	9	AG413529 Mus muscu
C 22	38.8	7.7	947	6	CD324699 AGENCOURT
23	38.8	7.7	1019	6	CD326862 AGENCOURT
24	38.8	7.7	1376	9	AG320971 Mus muscu

ALIGNMENTS

RESULT 1.
LOCUS CV262174/c

DEFINITION trichocarpa x Populus nigra cdna clone WS02018_N09 3', mRNA

ACCESSION CV262174.1
VERSION GI:52515149
KEYWORDS EST.
SOURCE Populus balsamifera subsp. trichocarpa x Populus nigra

ORGANISM Populus balsamifera subsp. trichocarpa x Populus nigra
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Malpighiales; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 342)

AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Mason,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.

TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries

JOURNAL Unpublished (2004)

COMMENT Contact: Joerg Bohlmann

Genome BC forest genomics program
University of British Columbia
UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3

Tel: 1-604-822-0282
Fax: 1-604-822-6097

Email: bohlmann@interchange.ubc.ca
Plate: WS02018 row: N column: 09

High quality sequence stop: 342.

FEATURES
source

1. 342
Location/Qualifiers
/organism="Populus balsamifera subsp. trichocarpa x Populus nigra"
/mol_type="mRNA"
/cultivar="Nxm6"
/db_xref="taxon:293756"
/clone="WS02018_N09"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/clone_lib="PTXN-IB-N-A-11"
/notes="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); Sapling trees

BI747328 rm38g08.y
BG787349 SEADUMC007
AL051744 Drosophil
AL274920 Tetraodon
BM168626 EST571149
AW932684 EST358527
BE353770 EST355113
AW934430 EST360273
AW933909 EST359752
AW037367 EST275905
CN384642 LE2TR01H2
CK264617 EST710695
AG304801 Mus muscu
Z444777 HSC2LE031 n
AQ417736 RPCI-11-1
BX461806 BX461806
BE052270 GA_Ea000
CR102355 Reverse s
AL104216 Drosophil
CB170882 IBC602600
AZ982725 2M0263E11


```

/clone="MSMg01-093D05.TJ"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      8.5%; Score 42.8; DB 9; Length 1279;
Best Local Similarity 51.6%; Pred. No. 1.2; Mismatches 0; Gaps 0;
Matches 98; Conservative 0; Indels 92; Indels 0; Gaps 0;

QY 304 GTTTTCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTATTTTATAGCCCTTGCGATA 363
DB 670 GTTATCTTTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 729
QY 364 TGTATATATAGCTGATGCACTGAGTGTGTTTACCTGATTTTCTTTTACAGCAAAATG 423
DB 730 CCCTCTTTTATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 789
QY 424 TGAGAGTTTATCTTCAATATATCATCCGTTTAAAGCATATTTTATTTTCTTATATCTG 483
DB 790 TCTTTTCTTTCTTCTCCATCTCTCTCTTTTATTTTATTTTCTTTTCTTTTCTTTTCTTT 849
QY 484 GCTTCGTTAC 493
DB 850 CTTTCCTTC 859

RESULT 4
CNS0039G      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL063921
VERSION      AL063921.1 GI:4941778
KEYWORDS
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN
Query Match      8.4%; Score 42.2; DB 9; Length 1101;

```

```

Best Local Similarity 12.5%; Pred. No. 1.7;
Matches 59; Conservative 223; Mismatches 189; Indels 0; Gaps 0;

QY 1 AATAGTTTAATTTCTGCTCGATACCTCTAGTGTTCATTTGATTATGCACTACTTTGCTG 60
DB 629 WHEWTTTAAAYYYTTCMYYYHYMMHHAHAHAAMWTTTHTWTHAYHATYHYYYMYC 688
QY 61 ACTATCTTTCTCAGGGGCTTTGGAGTACCACAAAGTCATAACATCCAGTCATCATGT 120
DB 689 ANMCMCTHTCHCYYYHYHTAHHTHTHHWYAHYMYMYWYMYCTACTYHYHHHH 748
QY 121 AAATCTCTCGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 749 YHWAYHTTWTAWAHAMMMHHAHAHAAMWTTTHTHHHTTTHYHHTTTHYHHTTTHY 808
QY 181 GGCAGCCATATCTCTGGAATCCAGAGGATATGTTAGCTAGTCAAGATTTGGGGGAAATA 240
DB 809 YMCYTHCHYHYHTAYTCWTHTHMMWTTTHTWHTTTHMMWTTTHTWHTTTHMMWTTT 868
QY 241 CCGGATGAATCAGAGAGGCTCCACACTGTTTAAAGCGAGAGTGGCAAACTTTGCG 300
DB 869 TWTWATHCWAACMTMHHHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 928
QY 301 CAAGTTTTCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTATTTTAGCCCTTC 360
DB 929 WMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 988
QY 361 ATATGCTTATATATAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 420
DB 989 HYHYMCTCHYCTHYCTHYCTHYCTHYCTHYCTHYCTHYCTHYCTHYCTHYCTHYCT 1048
QY 421 ATGTGAGAGTTTATTTCTTCAACTATATATATATATATATATATATATATATATATAT 471
DB 1049 ACTCHWTTWYHTCTWYHYHTYHMMWAWMMHMMHMMHMMHMMHMMHMMHMMHMMH 1099

RESULT 5
BU570551/c
LOCUS      BU570551      714 bp      mRNA      linear      EST 16-SEP-2002
DEFINITION      AGENCOURT_10405038 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6622324
5', mRNA sequence.
ACCESSION      BU570551
VERSION      BU570551.1 GI:22920851
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 714)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapbs@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2875 row: h column: 04
High quality sequence stop: 500.
Location/Qualifiers
1. .714
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6622324"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgcctcgcc); Site_2: SfiI (ggcattatggcc); 5' and

```

3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 8.3%; Score 41.8; DB 5; Length 714;
 Best Local Similarity 55.1%; Pred. No. 2;
 Matches 102; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 315 TTCAGTAAGTGGCCCTTTTTCACCTCCACCACTTATTTTACGCTTGCATATGCTTATATAT 374
 Db 260 TTCAATTATTTTCATCATTTTCATTCATCATTTTATTTTCATCATTTTCATTCATTTTCAT -C 202

QY 375 AGCTGATTCGCACTGAGTGTGTTTACCTGATTTTCTGTACAGCAATGTGAGAGTTTGA 434
 Db 201 ATTTCATTTTCATTTTCATATCATTTTCATTCATTTCTCTTTTAAATTTTCATTAATTTCA 142

QY 435 TTCTTCACTATATCATTCGTTTACGATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAC 494
 Db 141 TCATTCACCTTCATCATTTTCGTTTTCATTCATTTTCATTTTCGATGGAGACTCGCTCCATTACC 82

QY 495 AATGCC 499
 Db 81 CAGGC 77

RESULT 6

BE045401/c
 LOCUS
 DEFINITION
 h21f08.x1 NCI CGAP Lu24 Homo sapiens CDNA clone IMAGE:2955783 3' similar to SW:RL2B_HUMAN P29316 60S RIBOSOMAL PROTEIN L23A. ;, mRNA sequence.
 BE045401
 BE045401.1 GI:8362539
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 391)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-t@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cdna Library Preparation: M. Bento Soares, Ph.D.
 cdna Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: infoimage.llnl.gov
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers

FEATURES

source
 1. .391
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2955783"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu24"
 /note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized library NCI CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Ronaldo.

ORIGIN

Query Match 8.3%; Score 41.4; DB 2; Length 391;
 Best Local Similarity 52.3%; Pred. No. 2;
 Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 283 AAGATGGCAAACTTTTGGCAAGTTTTCATCTTTTCAGTAAGTGGCTTTTTCACCTCCAC 342
 Db 340 AATGTCACACAGCTTCTTCATAGTCTCTTCCATCCGGTATGTCGNTCTAACATCCAC 281

QY 343 AACTTATTTTACCTTGCATATGCTTATATATATAGTTCGCACTGATGTGTTACCTG 402
 Db 280 AACGAATRAACCGTGTCTCTCTCTATATCTCCCTATGCCACTTAGTGTGTCACCGG 221

QY 403 ATTTCTCTGTACAGCAATGTGAGAGTTTATTTCTTCAACTATATATCCATCC 454
 Db 220 ACAACGATGACACAGCAGTGTGTCAGCTTGTCTTCTACTGCCAGCGCTATCC 169

RESULT 7

CK520898
 LOCUS
 DEFINITION
 rswea0_004977.y1 swe Bombyx mori cDNA, mRNA sequence.
 CK520898
 CK520898.1 GI:40899756
 EST.
 Bombyx mori (domestic silkworm)
 Bombyx mori
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.

1 (bases 1 to 449)
 XIAO, Q., ZHOU, Z., LU, C., CHENG, D., DAI, F., LI, B., ZHAO, P., ZHAO, X., CHENG, T., CHAI, C., PAN, G., XU, J., LIU, C., LIN, Y., QIAN, J., HOU, Y., WU, Z., LI, G., PAN, M., LI, C., SHEN, Y., LAN, X., YUAN, L., LI, T., XU, H., YANG, G., WAN, Y., ZHU, Y., YU, M., SHEN, W., WU, D., XIANG, Z., YU, J., WANG, J., LI, R. Q., SHI, J. P., LI, H., LI, G. Y., SU, J. N., WANG, X. L., LI, Q. Q., ZHANG, Z. J., WU, Q. F., LI, J., ZHANG, Q. P., WEI, N., XU, J. Z., SUN, H. B., DONG, L., LIU, D. Y., ZHAO, S. L., ZHAO, X. L., MENG, Q. S., LAN, F. D., HUANG, X. G., LI, Y. Z., FANG, Y. Q., LI, C. F., LI, D. W., SUN, Y. Q., ZHANG, Z. P., YANG, Z., HUANG, Y. Q., XI, Y., QI, Q. H., HE, D. D., HUANG, H. Y., ZHANG, X. W., WANG, Z. Q., LI, W. J., CAO, Y. Z., WANG, J., YE, J., JI, H., LI, S. T., NI, P. X., ZHANG, J. G., ZHANG, Y., ZHENG, H. K., YE, C., WANG, J., WONG, G. K. S. and YANG, H. M.
 A draft sequence for the genome of the domesticated silkworm (Bombyx mori)

Unpublished (2004)

JOURNAL

Contact: Zhongguai Xiang
 Southwest Agricultural University
 Chongqing Beibei

COMMENT

Tel: 86-23-68251123

Fax: 86-23-68251128

Email: xzheswau.cq.cn

Location/Qualifiers

FEATURES

source
 1. .449
 /organism="Bombyx mori"
 /mol_type="mRNA"
 /strain="DaZhao (p50)"
 /db_xref="taxon:7091"
 /sex="female"
 /cell_type="Hemocyte"
 /dev_stage="5th-instar day-3 larva"
 /clone_lib="swe"
 /note="Vector: pBluescript II SK(+)"

ORIGIN

Query Match	8.2%;	Score 41.2;	DB 7;	Length 449;
Best Local Similarity	50.0%;	Fred.No. 2.6;		
Matches 103;	Conservative 0;	Mismatches 103;	Indels 0;	Gaps 0;
291	AAAACTTTGGCAAGTTTTGCATCTTTTCAGTAAGTGGCGCTTTTTCACCTCCACAACATTATT	350		
35	AAAAATTTTGTCAAGTTTATGTAATAATGATTAATAGATAATTTTACTTTCGGTGGCTATGAC	94		
351	TTAGCCTTGCATATGCTTTATATATAGCTGATTTGCAACTGTAGTGTGTACTGATTTCTCGT	410		
95	TTATTTTATGTGGATGCCATTATATTTTAAATTTTTTTCATTCGTAATTTACACCTGAAACT	154		
411	TTACAGCCAAATGTGAGAGTTTTTATCTTCCAACTATATCCATCGTTTTTAAAGCATATTTTA	470		
155	ACGTTGGCCAAACATCAGAATGTGTTTGGTTAATAATAATGCTCAGAAAATTCAAATTTAC	214		
471	TTTCTTATATCTCGCTTCGTTACCAA	496		
215	TAAATTTTATTTAGATTTAAATGGCAA	240		

RESULT 8
 CD681805/c
 LOCUS
 DEFINITION
 CD681805 587 bp mRNA linear EST 24-JUN-2003
 rj31d01.v1 Melloidogyne chitwoodi egg S11 TOPO v1 Melloidogyne
 chitwoodi cDNA 5' similar to TR:Q9TXU7 Q9TXU7 H06H21.3 PROTEIN. [1]
 ;, mRNA sequence.
 CD681805
 CD681805.1 GI:32183571
 EST.
 Melloidogyne chitwoodi
 Melloidogyne chitwoodi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Melloidogyninae; Melloidogyne.
 1 (bases 1 to 587)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
 Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
 Tsagarashvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
 Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

```

FEATURES
high quality sequence btop: 387.
location/Qualifiers
    1..587
        /organism="Meloidegynae chitwoodi"
        /mol_type="mRNA"
        /db_xref="taxon:59747"
        /dev_stage="eggs"
        /lab_host="DH10B"
        /clone_lib="Meloidegynae chitwoodi egg SL1 TOPO v1"
        /note="Vector: pCR11-TOPO (Invitrogen); Site 1: EcoRI;
        Site 2: EcoRI; The library was constructed by Claire

```

Murphy and Dr. Makedonka Dautova at Washington University, St. Louis. (dT)-SL1 PCR based library, MeloIdodogyne chitwoodi egg cDNA PCR products of size >400 nucleotides containing SL1 on the 5' end and oligo(dT) on the 3' end were non-directionally cloned into pCRII-TOPO(Invitrogen) following the ToPO TA cloning protocol. Eggs were provided by Dr. David Bird of North Carolina State University, Raleigh, NC (david.bird@ncsu.edu)."

ORIGIN	Query Match	8.2%	Score 41.2;	DB 6;	Length 587;
	Best Local Similarity	55.6%;	Pred. No. 2.7;		
	Matches 79;	Conservative 0;	Mismatches 63;	Indels 0;	Gaps 0;
Qy	343	AACTATTATTTACGCTTGCATATGCTTATATATATAGCTGATGCAACTGTAGTTGTTACCTG	402		
Db	580	AACTTTTATTATCCCTTACTAAATTTTAAATACAGCGAAATTTAAATAAATATATATATTTT	531		
Qy	403	ATTTCCTGTTACAGCCAAATGTGAGAGTTTATTCTTCAACTATATATCCATCCGTTTAAAGC	462		
Db	520	ATTTCCAATATATATCAATCTTAAATTACTATTATTAACCTTCTTCCCTTCGTTTTAAC	461		
Qy	463	ATATTTTATTTCCTTATATCTGG	484		
Db	460	ATTGTGTTCTCTTCAACATCG	439		

RESULT	9
LOCUS	CD6822973/c
DEFINITION	rj28g12.y1 Meloidogyne chitwoodi egg SL1 TOPO v1 Meloidogyne chitwoodi cDNA 5' similar to TR:Q9TXU7 Q9TXU7 H06H21.3 PROTEIN. [1]
ACCESSION	CD6822973
VERSION	CD6822973.1
KEYWORDS	GI:32184739
SOURCE	EST.
ORGANISM	Meloidogyne chitwoodi Meloidogyne chitwoodi Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne. 1 (bases 1 to 638)
REFERENCE	McCartner,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagariswili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. The Washington Univ. Nematode EST Project, 1999 Unpublished (1999) Contact: McCartner JP
TITLE	
JOURNAL	
COMMENT	

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

The library was constructed by Claire Murphy and Dr. Makedonka Dautova at Washington University, St. Louis
(mdautova@watson.wustl.edu). Oligo(dT)-SLI PCR based library.
Meloidegynine chitwoodi egg cDNA PCR products of size >400
nucleotides containing SLI on the 5' end and oligo(dT) on the 3'
end were non-directionally ally cloned into pCRII-TOPO(Invitrogen)
following the TOPO TA cloning protocol. Eggs were provided by Dr.
David Bird of North Carolina State University, Raleigh, NC
(david_bird@ncsu.edu).

Seq primer: SLI primer
High quality sequence stop: 638.
Location/Qualifiers
1..638
/organism="Meloidegynine chitwoodi"
/mol_type="mRNA"
/db_xref="taxon:59747"

FEATURES
Source

```

/dev_stage="eggs"
/lab_host="DH10B"
/clone_lib="Meloiodogyne chitwoodi egg SL1 TOPO v1"
/note="Vector: pCRII-TOPO (Invitrogen); Site 1: EcoRI;
Site 2: EcoRI; The library was constructed by Claire
Murphy and Dr. Makedonka Dautova at Washington University,
St. Louis. Oligo (dT)-SL1 PCR based library. Meloiodogyne
chitwoodi egg cDNA PCR products of size >400 nucleotides
containing SL1 on the 5' end and oligo(dT) on the 3' end
were non-directionally cloned into pCRII-TOPO (Invitrogen)
following the TOPO TA cloning protocol. Eggs were provided
by Dr. David Bird of North Carolina State University,
Raleigh, NC (david_bird@ncsu.edu)."

```

ORIGIN

```

Query Match      8.2%; Score 41.2; DB 6; Length 638;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 79; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 343 AACTTATTTTAGCTTGCATATGCTTATATATAGCTGATTCGCAACTGTAGTTGTTACCTG 402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 AACTTTTATTTACCTTACTTAATATTTAAATACAGCGAATTTAAATATAATATATATTT 572
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 403 ATTTCTGTTACAGCAATGAGAGTTTATCTTCAACTATATCATCGTTTAAAGC 462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 ATTTCCAAATATATCAATCTTAATTTACTTATTTAAACCTTTCTTCCCTCGTTTAAAC 512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 463 ATATTTTATTTCTTATATCTGG 484
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 ATTGTGTTCTTCTCAACATCG 490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10

```

CK310894/c
LOCUS      824 bp mRNA linear EST 01-MAR-2004
DEFINITION SB02007BD10.f1 normalized Keck-Tagu Library SB02 Taeniopygia
guttata cDNA clone SB02007BD10.f1 5, mRNA sequence.

```

ACCESSION

```

CK310894
CK310894.1 GI:4820468

```

KEYWORDS

SOURCE

ORGANISM

```

Taeniopygia guttata
Taeniopygia guttata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Estrilidae;
Estrilidae; Taeniopygia.

```

REFERENCE

AUTHORS

```

1 (bases 1 to 824)
Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
and Liu,L.

```

```

The Songbird Neurogenomics Initiative: An Evolving Public Resource
for Study of Genes, Brain, and Behavior
Unpublished (2004)

```

TITLE

JOURNAL

COMMENT

```

Contact: David P. Clayton
University of Illinois
BL07 CLSL, 601 S. Goodwin, Urbana, IL 61801, USA
Tel: 217 244 3668
Fax: 217 244 1648
Email: dclayton@uiuc.edu

```

```

Base Calling/Quality Scores: PHRED from Washington University
Genome Center.

```

```

Vector Trimming: Cross match from Washington University Genome
Center PHRAP suite. Low quality bases (Phred score < 20) were
trimmed from both ends of the sequence by an in-house script.
This sequence is vector free and at least 200 bp in length. Funded
by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'

```

```

PCR Primers
FORWARD: TAATACGACTCACTATAGG(T7)
BACKWARD: ATTAACCTCACTAAAG(T3)
Insert Length: 824 Std Error: 0.00
Plate: SB02007B1 row: D column: 10
Seq primer: TAATACGACTCACTATAGG (T7)
High quality sequence stop: 824.
Location/Qualifiers

```

FEATURES

source

```

1..824
/organism="Taeniopygia guttata"
/mol_type="mRNA"
/db_xref="taxon:59729"
/clone="SB02007BD10.f1"
/tissue_type="brain"
/dev_stage="late embryo, post-hatch days 1, 10, 20, 45,
and adult (pooled)"
/lab_host="DH10B"
/clone_lib="normalized Keck-Tagu Library SB02"
/note="Organ: brain; Vector: pBS II SK(+); Site 1:
EcoRI(5' side of insert); Site 2: NotI (3' side of
insert); The library was constructed and normalized as
described by Bonaldo, M.F., Lennon, G. and Soares, M.B.
(1996), Genome Research 6(9): 791-806. An identifying tag
was added at the 3' during cDNA synthesis:
insertAAAAAAAAAAAAAAAAAATGCGA."

```

ORIGIN

```

Query Match      8.1%; Score 40.6; DB 7; Length 824;
Best Local Similarity 50.8%; Pred. No. 4.3;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 292 AAACTTGCAAGTTTGCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACATATTT 351
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 542 AAACCTTTTCAAAATTTGTAATTTTAACTTTTAAATTTTGTCTTTTAACTTTTAACTT 483
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 352 TAGCCCTTGCAATGCTTATATATAGCTGATTCCAACTGTAGTTGTTACCTGATTCCTGT 411
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 TTAACTTTTAAATTTTAAATTTTAACTTTTAAATTTTGTCTTTTAACTTTTAACTCCT 423
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 412 TACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCATCGTTTAAAGCATATTTAT 471
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 AACTTTTAAATTTTAAATTTTAAATTTTAACTTTTAAATTTTAACTTTTAACTTTTCAA 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 472 TTCTTATATCT 482
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 ATTTTTAAATTT 352
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 11

AG390927

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

```

AG390927      1316 bp DNA linear GSS 03-JUN-2004
Mus musculus molossinus DNA, clone:MSMg01-207N11.T7, genomic survey
sequence.
AG390927      AG390927.1 GI:48009498
GSS.
Mus musculus molossinus
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
Unpublished
2 (bases 1 to 1316)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgpp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7

```

```
LIBRARY
Vector      : pBACE3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI
FEATURES
source
1. .1316
Location/Qualifiers
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSG01-207N11.77"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSG01 Mouse Male BAC Library"
ORIGIN
Query Match      8.1%; Score 40.4; DB 9; Length 1316;
Best Local Similarity 51.1%; Pred. No. 5.4;
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 305 TTTTGCATCTTTTCAAGTGGCCCTTTTCACTCCCAACTATTTTACCTTGCATAT 364
    |||||
Db 735 TTTTCTATCTTTTATTTCTTATCTTGACTTATATTTTAAATTTATCTCTTTAT 794
    |||||
QY 365 GCATTATATAGCTGATGCACTGATGTTGTTACCTGATTCCTGTCAGCAAAATG 424
    |||||
Db 795 ACTTTTATTTATCTTTTGTGACCTCTCTTTCTTTTCTTTTCTTTTACCCAAACT 854
    |||||
QY 425 GAGAGTTTATCTTCAACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATCTGG 484
    |||||
Db 855 TACTCTTATTTCTATACATTTCTTTTATTTATCTTCTTATCTTCTTACTTTATAC 914
    |||||
QY 485 CTTCGT 490
    |||
Db 915 TTTCTT 920

RESULT 12
AF173385
LOCUS      Homo sapiens MSTP101 (MST101) mRNA, complete cds.
DEFINITION
ACCESSION  AF173385
VERSION     AF173385.1 GI:33338007
KEYWORDS   HTc.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 816)
AUTHORS    Zhao,B., Tong,Y.K., Xu,H.S., Qin,B.M., Liu,B., Wang,X.Y., Zhang,Q.,
Song,L., Gao,Y., Zhang,C.L., Ye,J., Ji,X.J., Liu,B.H., Lu,H.,
Chen,J.Z., Cai,M.Q., Zheng,W.Y., Teng,C.Y., Liu,Q., Yu,L.T.,
Lin,J., Gong,Q., Zhang,A.M., Gao,R.L. and Hui,R.T.
Homo sapiens normal aorta mRNA MST101
TITLE      Homo sapiens normal aorta mRNA MST101
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 816)
Zhang,B., Tong,Y.K., Xu,H.S., Qin,B.M., Liu,B., Wang,X.Y., Zhang,Q.,
Song,L., Gao,Y., Zhang,C.L., Ye,J., Ji,X.J., Liu,B.H., Lu,H.,
Chen,J.Z., Cai,M.Q., Zheng,W.Y., Teng,C.Y., Liu,Q., Yu,L.T.,
Lin,J., Gong,Q., Zhang,A.M., Gao,R.L. and Hui,R.T.
Direct Submission
Submitted (30-JUL-1999) Molecular Medicine Center for
Cardiovascular Disease, Cardiovascular Institute, CAMS & PUMC, 167
Bei Li Shi Lu, Beijing 100037, P.R. China
FEATURES
source
1. .816
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="aorta"
1. .816
/gene="MST101"
632. .781

/gene="MST101"
/codon_start=1
/product="MSTP101"
/protein_id="AAQ13626.1"
/db_xref="GI:33338008"
/translation="MNGENMRKPDGPVSTFLRHFI PDGAVLLIILKLFPSGNTDLF
FKPER"

ORIGIN
Query Match      8.0%; Score 40; DB 3; Length 816;
Best Local Similarity 53.1%; Pred. No. 6.3;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 321 AAGTGGCCTTTTTCACCTCCCAACTATTTTACCTTGCATATGCTTATATATAGTGA 380
    |||||
Db 81 AAGTGGCCCGCCGCCCAACAAATTTGGATCTTTTCTCTTTTAATAAGCAGG 140
    |||||
QY 381 TTGCAACTGTAGTGTGTTACCTGATTTCTGTTTACAGCAATGTGAGAGTTTATTTCTTC 440
    |||||
Db 141 GCTGTATTATCTTGAATATATGTTGCTTCTTGAACATATTAAGATATATTTATTTGCC 200
    |||||
QY 441 AACTATATCCATCCGTTTAAAGCATATTTTATTTCTTATAT 480
    |||||
Db 201 ACTAGTATTAAATTTTGTCACATTTTCATTAATACAT 240

RESULT 13
AG305042
LOCUS      Mus musculus molossinus DNA, clone:MSG01-085H11.TJ, genomic survey
DEFINITION
ACCESSION  AG305042
VERSION     AG305042.1 GI:47877996
KEYWORDS   GSS.
SOURCE     Mus musculus molossinus
ORGANISM   Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSG01
Unpublished
2 (bases 1 to 1188)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan 305-0874
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSG01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center, Chemical Research (RIKEN) 3-1-1
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector      : pBACE3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI
FEATURES
source
1. .1188
Location/Qualifiers
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSG01-085H11.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSG01 Mouse Male BAC Library"
```

ORIGIN

Query Match 7.9%; Score 39.8; DB 9; Length 1188;
 Best Local Similarity 50.8%; Pred. No. 7.8;
 Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 305 TTTTGCATCTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTATTTTAGCCCTTGCAATAT 364
 |||||
 Db 815 TTTTATATCTTCTTTTGAATTTATTTCTCTTTTCTCTTTTCTCTTTTCTCTTTTCTCTTT 874
 |||||
 QY 365 GCTTATATAGCTGATGCAACTGAGTGTGTTACCTGATTTCTCTGTTACAGCAATGCT 424
 |||||
 Db 875 TCTAATTTATTTTATTAACCTCTTTTATTAATTTATTTTATTTTCTTATTTTACTTC 934
 |||||
 QY 425 GAGAGTTTATTTTCTCAACTATATCATCGTTTAAAGCATATTTTATTTCTTATATCTCG 484
 |||||
 Db 935 CTTTATTTATTTCTTTTATTTCTTCTGCTTTCTTTCTTTTATTTATTTCTTTTCTTCTGT 994
 |||||
 QY 485 CTTCTGTT 491
 |||||
 Db 995 CTACTTT 1001

RESULT 14

LOCUS CL647966
 DEFINITION CH213-157E13.17 CH213 Gasterosteus aculeatus genomic clone
 CH213-157E13.5', genomic survey sequence.

ACCESSION CL647966
 VERSION CL647966.1 GI:49667390
 KEYWORDS GSS.
 SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
 Gasterosteidae; Gasterosteus.
 REFERENCE 1 (bases 1 to 1476)
 AUTHORS Kingsley, D., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M.
 TITLE Expressed sequence tags from Gasterosteus aculeatus
 JOURNAL Unpublished (2004)
 COMMENT Contact: Grimwood, Jane
 Stanford Human Genome Center
 Stanford University School of Medicine
 975 S California Avenue, Palo Alto, CA 94304, USA
 Tel: 650 320 5917
 Fax: 650 320 5801
 Email: jane@shgc.stanford.edu
 Plate: 157
 Class: BAC ends
 High quality sequence start: 22
 High quality sequence stop: 864.

FEATURES

source
 1..1476
 Location/Qualifiers
 /organism="Gasterosteus aculeatus"
 /mol_type="genomic DNA"
 /strain="Salmon River"
 /db_xref="taxon:69293"
 /clone="CH213-157E13"
 /sex="Mixed"
 /cell_type="Blood"
 /clone_lib="CH213"
 /notes="Vector: pTARBAC2.1; Site 1: EcoRI; The sequence of the clone was established as a mapping and sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (<http://cegs.stanford.edu>). The clone was isolated from the BAC library CHORI-213 built by Pieter deJong in collaboration with the Stanford Genome Evolution Center (<http://www.chori.org/bacpac/>). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm)."

ORIGIN

Query Match 7.9%; Score 39.8; DB 9; Length 1476;
 Best Local Similarity 50.8%; Pred. No. 8.1;
 Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 290 GAAACTTTGCCAGTTTGCATCTTTTCAGTAAGTGGCTTTTTCACCTCCACAACTTAT 349
 |||||
 Db 1028 CAAAATGTGGCTTCGATTTGATTTTACTCTTTTAAATGATATCTTTCTCTATCTTTT 1087
 |||||
 QY 350 TTTAGCCTTGCAATATGCTTATATATAGCTGATTTGCAACTGATTTGTTTACCTGATTTCT 409
 |||||
 Db 1088 TCTATTTTCTCATATTTTCTTCTATATATTTTCTTACGTTTATCATTTCTT 1147
 |||||
 QY 410 GTTACAGCCAAATGTGAGAGTTTATTTCTTCAACTATATCCATCCGTTTAAAGCATATTT 469
 |||||
 Db 1148 TATCAATATTTTCTTCTTTTCTTTCTTATATCTTTTCTTATATCTTTTATTTCA 1207
 |||||
 QY 470 ATTTCCT 476
 |||||
 Db 1208 TTTTCCT 1214

RESULT 15
 LOCUS BU357425
 DEFINITION 603473713F1 CSEQCHN70 Gallus gallus cDNA clone ChEST354i3 5', mRNA
 sequence.

ACCESSION BU357425
 VERSION BU357425
 KEYWORDS BU357425.1 GI:25865426
 SOURCE EST.
 ORGANISM Gallus gallus (chicken)
 Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 774)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..774
 Location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="ChEST354i3"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN70"
 /note="Organ: hearts; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2534.52 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929C-1_COPY_9829_10329
Perfect score: 501
Sequence: 1 gggctctagttattgattt.....gtgatttgaaatgaacaatt 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	10329	6	AX063582	AX063582 Sequence
2	501	100.0	10329	8	AF213628	Arabidops
C 3	501	100.0	120965	5	AC026875	Genomic s
4	224.4	44.8	6571	6	AX063583	AX063583 Sequence
5	219	43.7	6554	8	AF213627	Arabidops
6	178	35.5	939	8	AK117959	Arabidops
C 7	61.2	12.2	223966	2	CR318666	CR318666 Danio rer
C 8	57.8	11.5	5690	6	AX277910	AX277910 Sequence
C 9	57.8	11.5	5690	6	AX323591	AX323591 Sequence
C 10	57	11.4	5153	6	AX345857	AX345857 Sequence
C 11	57	11.4	10329	6	AX347025	AX347025 Sequence
12	57	11.4	250022	3	AE014824	AE014824 Plasmodiu
13	56	11.2	348034	3	CR382400	CR382400 Plasmodiu
C 14	55.8	11.1	110000	2	BX901979	Continuation (2 of
C 15	55.6	11.1	205429	2	AC005506	AC005506 Plasmodiu
C 16	55.6	11.1	253132	3	PFMAL3P7	AE014846 Plasmodiu
C 17	55.2	11.0	253305	3	PFMAL3P7	AE014559 Plasmodiu
C 18	55	11.0	250029	3	AE014838	AE014838 Plasmodiu
19	54.6	10.9	244097	2	CR354394	CR354394 Danio rer

C 20	54.4	10.9	254050	3	PPA29358	AL929358 Plasmodiu
21	54.2	10.8	250707	3	AE014848	AE014848 Plasmodiu
C 22	54	10.8	202863	2	EX927344	EX927344 Danio rer
23	53.8	10.7	176034	2	CR524826	CR524826 Danio rer
24	53.8	10.7	178095	2	CR753874	CR753874 Danio rer
C 25	53.6	10.7	18512	6	AX345879	AX345879 Sequence
26	53.4	10.7	146917	2	CR450801	CR450801 Danio rer
27	53.4	10.7	173419	2	EX908733	EX908733 Danio rer
C 28	53.2	10.6	40361	9	EX682235	EX682235 Human DNA
C 29	53.2	10.6	349980	6	AX344564	AX344564 Sequence
30	53	10.6	67970	3	PFMAL1P3	AL031746 Plasmodiu
C 31	53	10.6	212881	2	CR450705	CR450705 Danio rer
32	52.8	10.5	84710	5	AL928945	AL928945 Zebrafish
C 33	52.6	10.5	250029	3	AE014820	AE014820 Plasmodiu
C 34	52.6	10.5	349751	3	PFMAL4P3	AL034576 Plasmodiu
35	52.4	10.5	158053	2	CR293506	CR293506 Danio rer
36	52.4	10.5	198047	2	CR376763	CR376763 Danio rer
C 37	52	10.4	152512	10	AC131766	AC131766 Mus muscu
38	52	10.4	159835	2	CR387984	CR387984 Danio rer
39	52	10.4	160795	2	CR848031	CR848031 Danio rer
C 40	52	10.4	178273	2	AC005308	AC005308 Plasmodiu
41	52	10.4	215584	10	AC125215	AC125215 Mus muscu
42	52	10.4	223662	2	CR769783	CR769783 Danio rer
43	52	10.4	250531	3	AE014845	AE014845 Plasmodiu
C 44	52	10.4	349980	6	AX344572	AX344572 Sequence
C 45	51.8	10.3	110000	2	PFMAL7P1_04	Continuation (5 of

ALIGNMENTS

RESULT 1
AX063582
LOCUS AX063582 10329 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100801.
ACCESSION AX063582
VERSION AX063582.1 GI:12541306
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE 1
AUTHORS Habu,Y., Mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 1 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft m.b.H. (AT)

FEATURES
source 1.10329
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

Query Match 100.0%; Score 501; DB 6; Length 10329;
Best Local Similarity 100.0%; Pred. No. 2.1e-84;

[illegible]

mRNA

[illegible]

3113,

gb|A1812185.1, emb|z17990.1, gb|AW267999.1, gb|N96009.1,
db|JAV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"
/translation="MGKEFHINIVVIGHVSGSKSTTGHLYIKLGGIDKRVIERFEK
EAEWMKRSFYAWLDLKAEREGITIDIALMKPFTTKYCTVIDAPGRHDFIKNM
TGTQADCAVLIIIDSTGGFEAGLSKDSQTEHALLAFTLGVKQMICCNKWDATTP
KTSKARYDIIKEVSYSLLKGVNPDKIPIFPISEFGNMIERSTNLNDYKGPITLLE
ALDQINERKPSDKPLRLQDVIYKIGGIVPVGKVTGMIKPGMVVTFAPGLUTE
VNSVEMHESLEALPGDNVFNKNVAVKDLKRGYVASNKDDPAKGAANFTSQVII
MKHPGQINGYAPVLDDTSHIAVKFSEILTKIDRRSGKEIEKEPKFLKNGDAGVNM
TPTKPMVVFTESEYPLGRFAVDRMQTVAVGVKISVDKDDPTGAKVTKAAVKKGAK"
complement (20787..20857)
/products="trNA-Lys"
/note="codon recognized: AAA; putative trNA-Lys"
/evidence=not experimental
complement (join(21956..22189,22535..22600,22713..22929,
23041..23126))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translation="MCLNPLPSATPIPRAEEREECSMGFPNHOILLISFLCKRPA
IIMKGGGGGGGGSGPTAAAAAQKQKALLQRTVETDITSVDNFTQINVS RV
SDPPVKNSSQTYMENRASNMYQAADSLIKLVSELKQTAIFSGFASLNDHVRQRIEF
DOEAENRLLAIADNASANIKLESHYSQAQRLTDI"
join(23576..23689,24080..24258,24347..24522,24635..24726)
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translation="MTGLGARLVAPMIILLFIPIELVKAETVITLPETPSDKLALSN
KQALQNGCISMSCFYSLSDTTRCLLSPCLLILOIKEDTAMFVKFCVPCWKCKKLG
NWDGLGKAWEGDEIEVGEVDCGTSRACVTKVHSYTFMLFYNGEVSFKYKGRD
VESLKAFVVEETEKAABKAQEDKEL"
complement (join(25423..25563,25675..26048,26127..26330,
26585..26783,26941..27842,27936..27991,28208..28413))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translation="MDSVRRKINNTTQSSPKSPKFSVYRNPALAAASTANSIRP
SKVLFIILLSTASAFSLISPMAGEKLTNALTFGKISQEAAYTVTKAWOGLVTLFC
IGAMNALSIGISHRAKPAARGETKSKETKDOESLSNROLELIGIKKADQSVSEF
PKSRPAKPAQSLPGLPIHQTLTSAHKSQGGKDLNRNGSLISSFTSKQMS
PSMYLPSSPVSNNLSQDKAVSPWSGRRSSAKDITTEEQQLLAIDEKITE
SAGKMTPLRTVGSFAMASPTVGGTSGAGATSTPLRPMSPGQAFPTTPPKVQ
GDFPNMSLERAIEGFIHLGVPOIEDMDRLQWCSVLLKPLNKVETSHIQVMT
ASKLVNVTVSQVSDLPNTGATATL PVDRTKGWQPSVSLDEDALLHOLRANLVQAI
DASMQKLRTEQFOQQOQQOQQAALI PMQECVDAISEHRLQGLMKGEWVKGLIP
RSIPADYTVQRILAAGTCVKNYINGGADTRKNKWSQPTFDSHLLYLFCAF
LEHDKWMLHDPSSYTGQASKNPLFLVLPKPERPEKFIYAVVSGVSTLHPGACVL
LBDKQSPPTFALYWKVQFTLQGTALWDSMLLI CHRIKVGYGGVVGRMGLSSALN
ILQVDSDDDD"
complement (join(29418..29491,29579..29620,29858..29942,
30386..30805))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translation="WSSSKPKKKARSDVVVNKASGRSKRSRSGSRTKTKSNKYNIVK
KXPEIYEI8ESSSSDSVBEAIRGDEAKNGSVVSKRNGKSGVIGIPTKTSKNREDDGG

AEADKIKFPMNRIRRMKSDNSAPOIMQDAVFLVNKATEMFLERSEBAYDSSVKDKK
KFIHYKHLSSVVSNDQRYEFLADSVPEKKAEEALEEWEWGMDTAG"
Query Match 100.0%; Score 501; DB 8; Length 120965;
Best Local Similarity 100.0%; Pred. No. 1.3e-84;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGTCTTATGTTATGATTTTATAGAGAGTTAAATAAGTATATATATATATATAGTAG 60
Db 57855 GGGTCTTATGTTATGATTTTATAGAGAGTTAAATAAGTATATATATATATATAGTAG 57796
QY 61 GTTACTTAATCTCTGTCGTTAATCTAATTTAGTGTAGTCAAGAACCGACTCGTTGGCTAAG 120
Db 57795 GTTACTTAATCTCTGTCGTTAATCTAATTTAGTGTAGTCAAGAACCGACTCGTTGGCTAAG 57736
QY 121 GTCTCTCTCTTTTGAACGCAACGTTCTACTTTTCATGTATATAAATACAGTCTGTATCACAC 180
Db 57735 GTCTCTCTCTTTTGAACGCAACGTTCTACTTTTCATGTATATAAATACAGTCTGTATCACAC 57676
QY 181 AACACAAATTTGATGATTGAAATATCTACTGATTTAACTTTATAGAAAAACCAATATATAG 240
Db 57675 AACACAAATTTGATGATTGAAATATCTACTGATTTAACTTTATAGAAAAACCAATATATAG 57616
QY 241 ACGGACAACTTTATAAACATGTCAAACTTCGAAGTTAAATTTAGACCCCAATATTTTA 300
Db 57615 ACGGACAACTTTATAAACATGTCAAACTTCGAAGTTAAATTTAGACCCCAATATTTTA 57556
QY 301 CAATTATAGATTTTAAATCTCAACTATTTTCTGTATGTTTAAAGAGATATCCGAGTCTTT 360
Db 57555 CAATTATAGATTTTAAATCTCAACTATTTTCTGTATGTTTAAAGAGATATCCGAGTCTTT 57496
QY 361 TCTTTCCAGTTTCCCACCGTCCCATGACTCCCCAGCAGTAGAAAAAGCCAAAAAGT 420
Db 57495 TCTTTCCAGTTTCCCACCGTCCCATGACTCCCCAGCAGTAGAAAAAGCCAAAAAGT 57436
QY 421 AAACAAAAAGTCGTTAAAAAGTTAAATTTAAAAAATATAGATTTGACGTTTACTTAA 480
Db 57435 AAACAAAAAGTCGTTAAAAAGTTAAATTTAAAAAATATAGATTTGACGTTTACTTAA 57376
QY 481 AGTGAATTTGAATTTGAACAATT 501
Db 57375 AGTGAATTTGAATTTGAACAATT 57355
RESULT 4
AX063583 LOCUS 6571 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 2 from Patent WO0100801.
ACCESSION AX063583
VERSION AX063583.1 GI:12541307
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1
AUTHORS Habu,Y., Mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 2 04-JAN-2001;
Novartis (AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
Location/Qualifiers
source 1..6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310..6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
/translation="MKDEKIGLTGRTIYTRSLAASIPASVQETPGLRRSSRGTPST

KVI TPASATRKSERLAPASPVKSGGIVYNSTPSSLRNRNGKTEVSLQSGKSDN
 SIRKGDSTPIDEQKDSVBESTDKIPIMARSARYALFRGKJKESEALVDSANBEEL
 VVVGSRNIPAGNDVOGKTDPPADAGSKRLPVDETSLDKGTDFPLKSVTEKIV
 LDASPIVETGDSVIGSPSENLETQKLDGKTDCSPANAESKTLPVGETSLAEKYQ
 KFODNNTCLPANAESKRLPVGETSLEKOTDPLKSTTETGKMWLASPIVETRDS
 VICSPTNLETKLVLKSTGLETDIVLPLKRDPTAEILDACATVANGDDHVMSSDG
 VIPSQCKNDNRPKMCVNCTCKKQKYNQDCQNRSCSVCIPQVEESNVYQDKMETGPP
 VTSREBENGQIKHGKSDSPKSVYPIETVPVQLSDVQLEQCTFLFKSLSLSL
 SKIDLALBETANSVKTCDHPYVMDASLKQTLTKNLEHILDIIVEKASGKHLLDKI
 MLTHIKNGULKAVFYQATQPTGELLGNILEDFVQFGFKSYEHGIVSGKNSAIN
 NFNKESCQVLLLETRACSTQIKLLRADAPILFGSSLNPSHDVHKVKEIKESCERT
 KIPRLXSVCTVEEAKLILARONNRONKAVENLRSLTHALLMGWASVLPKLDHFMSS
 ETPDGSVPQSIMDGVIEHFESSILSSKSGEENEVKLLLEAKHAGQTVSSBSTLEG
 EDHILKSDSPNIPFWSLKLGKNPMWKYPSDTPQNRKRQVYPEGSEASPKTGDDGN
 AKRKKASDDETPDPRVTDVDDPDKERKASGDHMKLASPEKVITLQSSASSTGDTL
 GDNDAFOLGSMGSHIGSIEEEMATTQFQIALPDESQRRLHTVLKPMKALCQVHLHSD
 ACTSMVGNFLEYIEHNRHIVPEEPALTSQWALISALVTAALLVKLISHKSELVRANSE
 LAFKSRVEVDYIVSILSCWKSFLFHTQGLQFDCFTGNSQSVSTPLKVNESLSGAT
 VRDEKINTKSMRNSDEBECKTEKRSCHYSTATRDIEKTSIGIKKKYKQVQKLVQBH
 BEKQWELNNAKQKLETYSKSVAAEIVITCSTSTVQGDGLLDHNYTERKPFDEIK
 SEKNBCLKLEQHEVAKLADBEACWNRKISWAKLKVCYPIOSGNNGHVPSSGN
 ISQWAPDQVICNNAVEATYADTNCAKSNQVPEAEKTLGTMSGASTQOVHEWMDVR
 NDETMDSVLSREOLTYSQSNHASITVPELIPADQCEBEAALNVHLSDEQNCDRIT
 SAASDEDSVRPEVLOSLENHASPEFSLNREALVTENRRTSHVGFETDNLLOQ
 NREDCSLQELIDELAMPVOHLASVETRGAESDQYQDIPCMPSSLAGKQDPPAAN
 TSENLBEALPEQASGETSDYAVIDQETWGAQDACLPSGSGVGTOSDILGANTEQNVTV
 NTAEPHBEAPVADGESVDYAVIDQETWGAQDACLPSGSGVGTOSDILGANTEQNVTV
 AOLPTDGSDAVVTGGSVPDQCADSPMLSPSGPNHPDPAVNI ELDGNTSVAEPHIS
 GSDACEMEISEPQOVERSTFANL FHEGGVHEHSAGVTALVPSILNNGTEOIAVQVPQ
 IIPPVENDPLHELEKLRRSESNKKTFEELKILKAEERKMAEVAQAEFRKFEHVE
 AERNRTTKLEKNNVLIMNKLAPLAKTCDKTKYSPSGAPRGKTIQLAQRAQVSA
 LRNYITAPQIQKASFPALVAPLQIQSSFPAGPAPLOQASFPFSSRPSRPSALL
 LNFVACPMQPROPLINASTPTSVTPATNPLGRSLPHLNSYRSPSSSTPVPATPTS
 SVPEQALYSAGVISOQOEOPOOSLSGSGOSNNNEVVCLDPS

[illegible]

RESULT 5	AF2113627	6554 bp	mRNA	linear	PLN 01-JUN-2000			
LOCUS	AF2113627							
DEFINITION	Arabidopsis thaliana MOM mRNA, complete cds.							
ACCESSION	AF2113627							
VERSION	AF2113627.1	GI:8132767						
KEYWORDS	Arabidopsis thaliana (thale cress)							
SOURCE	Arabidopsis thaliana							
ORGANISM	Arabidopsis thaliana							
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.							
REFERENCE	1 (bases 1 to 6554)							
AUTHORS	Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.							
TITLE	Disruption of the plant gene MOM releases transcriptional silencing of methylated genes							

JOURNAL Nature 405 (6783), 203-206 (2000)
 MEDLINE 20279299
 PUBMED 10821279
 REFERENCE 2 (bases 1 to 6554)
 AUTHORS Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland
 FEATURES Location/Qualifiers
 source 1...6554
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /cultivar="Zhuerich"
 /db_xref="taxon:3702"
 /chromosomes="1"
 /map="between ATEAT1 and phyA"
 CDS 310..6315
 /function="involved in the maintenance of transcriptional gene silencing"
 /codon_start=1
 /product="MOM"
 /protein_id="AAF73380.1"
 /db_xref="GI:8132768"
 /translation="MKDKDGLGTGTYITRSLAASIPASVEQETPLRRSRGRGTST
 KIRTPASATKGERLSPASVSKSGGIVKNSPTSSLRNRKTEKTEVLSQSGNNDL
 SVTKGDTSPDIQRKDSVEESTDKIPIINSAYSYALRGLKESALVDASPNREEL
 VVVGSRRIIPAGNDVVOGKTDCPPADAGSKRLPVDETSLDKGTDFPLKSVTETKIV
 LQASPIETGDSVIGSPSENLETKLQDKGTDCSPPAANAEKTLPLVGETSLEKEITPQ
 KFQDNTDCDLPANAEKRLPVGETSLEKTDPLKSTTETGKMLVASPIVETRDS
 VTCSPNTLETKLQKLVKTGLTDIVLPLKRRDIAEIELDACATVANGDDHVMGSDG
 VTPSPGCKNDNRPEMCTCKRQKVGDCNRVSCSIVQPVESDNTVTDQMKETGP
 VTSREYENGQIQHGKSSDPKFYSVYPRYVQLSDVQLSQYQOTLFSKLSLSLL
 SKIDGLABETLNVKRTCDHPYVMDASLKQLLTKNLEHLDEIVEIKASGKLHLDD
 MLTHIKKGLKAVFYQATQTPGELLGLNLSDFVQRPFGSKSYEHIKSSKNSAID
 NFNKFSQCVLLLETRACSTQTLKLRADAFILFGLSINLSPHDVHKVETIKIESCET
 KIFRYLSVCTVEKAILARQNNRQKNAVENLSRSLTHALLMWGASYLPDKLHFHSS
 EDTPDSVSEQIQMDGIVHEFSSILSSKGBENEVKLCILLBAKAAQGYTSSDSTLFG
 EPHIKLSDEESPNIFSKLIGKSNPMKYPSDTPQNRKRVOYFGSSEASPKTGDDGN
 AKRKLSDSDVTDPRVTDPPDDDERKASGKOHMGALSPKVTITLQSSCKSSGTGTL
 DGNDAFLYSGMHSIIGIPEDLMSADQWGIPIESQWRHLTVLKPMAKLCQVJHLS
 ACTAPVLYNFMETVIEHRIYEPATTFQAQIALSWIAALVQLILSHKESLVIRANSE
 LAFKPSRBEVDVYISLSCMSKLFLEHTQGLQFCFGTNSQSVSTKLNVESLGSAT
 VRDEKINTKSMENSDECMTEKRCSHYSTATRDIEKTIISGIKKKYKQVOKLQVEH
 BEKMELLNMYADKQKLETSKSVZAAVIRITCSRTSTQVGLDKLLDHYERKPFBIK
 ISQNECKLEQMHQVAKKLEADSEACINRIKSWAAKLKVCVPIQSGNNKHFGSSGN
 I SQNAPDVOI CKNANVEATYADTNCWAKNQVPAENTLGTMSGGTYQVQHMYDVR
 NDETMDVSALSRQLTKSQSNESHATVPEILIIPADCOBEFALANVHLSQDNCDRIT
 SAASDEDSVRPVEYQSLENISASPEFSLNREBALVTTENRRTSHVGDTDNILDQO
 NREDCSLDOEIPDELAMPVQHLASVVEITRGAESQYQGDICPMPSLQAGKQDPAA
 TSESNLEAIEFQSGASGVETITDFAASHQGDQVTCPLLSSPTGNQPAPEANIEQNTV
 NTSAPFHVAGDASVSGDYAVIDQETMGADQACSLPSGVSQDGLANIEQNTVTV
 QGLDTSADVATVSGPVSQDQADASPMPLSPGNNHPTAVNI EBLGNTSVABEHSI
 AGDCMETSEBGPQVRSFTFANLPHGGVHSAGVTALVPSLLNNGEQIAVQPVPO
 IIPPVNDPDLHELEKLRSESNKKTPEEKKSLKAELEKMAEVOAEFRKPFHEVE
 AEHNRTHQIKEDKNLIVNKNLAFSLKCTDCKVSPSGAPRGKIIQQLAQRQAQVSA
 LRNYTAPQQLASPPALVASPQLQSSFPAGPAPLPQAPQSPSSFSSVSRPRLALL
 LNFVCPMPQPRQIISNIAPTFPTPATNPLGRSPAPHLNSRSPSSSTPTVATATPT
 SVYPQALTVSAVSIQQQEQBPQQQLSSGLQSNNEVVCILSDDE"

[illegible]

Db 4596 AAAAAATCCCTCACTCTAANACTTAATATACAAAACTTTTATATAAATAAACCTTTCAAT 4537
QY 342 AAGAAGTATCCGAGTCTTTTCTTCCAGTTTCCCGACCGTCCCATGCTCCCGCCAG 401
Db 4536 TAAAAAATACTATTTCTCTTCATCTTCTACCTTCCCTATATAAATCCGATAAA 4477
QY 402 TAGAAAAAGCCAAAAAGTAACAAAAAGTCGTTAAAAAGTTTAAATTAATAAAAAA 458
Db 4476 CAAAAAATACTCAAAACTAAACTCAAAATAAAAAACGAAAAATTAATAATAATAA 4420

RESULT 9
AX323591/c
LOCUS AX323591 5690 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 79 from Patent WO0192565.
ACCESSION AX323591
VERSION AX323591.1 GI:18094339
SOURCE .
synthetic construct
synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A.; Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with dna transcription
JOURNAL Patent: WO 0192565-A 79 06-DEC-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
1..5690
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 11.5%; Score 57.8; DB 6; Length 5690;
Best Local Similarity 49.2%; Pred. No. 0.23;
Matches 146; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
QY 162 AATACAGTCTGATCACAACAATAATGATGATTGAAAAATACTACTGATTTAACTTTA 221
Db 4716 ACGAAACCGGAACCACTAAATAATAATAATAACATACATACATAATAACGAACCTA 4657
QY 222 TAGAAAAACCAATTTATAGCGACAACTTTTATAACATGTCAACTTCGAAGTTAAAT 281
Db 4656 ACGAAAAATATTAAACGAAACCTTAANTANTTTCAAACTATCCAACTTCGACTATNAAT 4597
QY 282 TTAAGACCCCATATTTTACAAATATAGATTTTAACTACCCCACTTTTGTGATCTTAA 341
Db 4596 AAAAAATCCCTCACTCTAANACTTAATATACAAAACTTTTATAAATAAACCTTTCAAT 4537
QY 342 AAGAAGTATCCGAGTCTTTTCTTCCAGTTTCCCGACCGTCCCATGCTCCCGCCAG 401
Db 4536 TAAAAAATACTATTTCTCTTCATCTTCTACCTTCCCTATATAAATCCGATAAA 4477
QY 402 TAGAAAAAGCCAAAAAGTAACAAAAAGTCGTTAAAAAGTTTAAATTAATAAAAAA 458
Db 4476 CAAAAAATACTCAAAACTAAACTCAAAATAAAAAACGAAAAATTAATAATAATAA 4420

RESULT 10
AX345857/c
LOCUS AX345857 5153 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 928 from Patent WO0200928.
ACCESSION AX345857
VERSION AX345857.1 GI:18493743
SOURCE .
synthetic construct
synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A.; Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 928 03-JAN-2002;

Epigenomics AG (DE)
Location/Qualifiers
1..5153
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 11.4%; Score 57; DB 6; Length 5153;
Best Local Similarity 48.9%; Pred. No. 0.33;
Matches 153; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 158 ATATAAATACAGTCTGATCACAACAATAATTTGATGTTGAAAAATCTACTGATTAAAC 217
Db 4931 ACAAAAATATAAATCAATAAAAAACAAAAACCTCAAAAAATAATACCATATCTACA 4872
QY 218 TTTATAGAAAACCCAAATTTATAGCGCAACTTTTATAACATGTCNAACTTTCGAAGTTA 277
Db 4871 ACTATCTAATCTTTTAAACAACCTTAAACAAAAACAAAAATAAAAAAATTCCTATATTA 4812
QY 278 AAATTTTAAGACCCCATAAATTTTACAATTTATAGATTTTAACTCTCAACTTATTGTGATG 337
Db 4811 ATATAATATCTAAAAAAACTTAACCTATATATAAAAAATAAACTAAATTCCTTCC 4752
QY 338 TTAAGAAGATATCCGAGTCTTTTCTTCCAGTTTCCCGACCGTCCCATGACTCCCGCCAG 397
Db 4751 TTACACCTTATATAAAAAATTAATTCAAAAATAATTAATACTTACATATTAACCTAAAA 4692
QY 398 CCAAGTAGAAAAGCCAAAAAGTAACAAAAAGTCGTTAAAAAGTTTAAATTAATAAAAAA 457
Db 4691 CCATAAAACCCCTAAAAAATAAACTTTAAACAATACCATTTCAAAAAACATAAAACAAA 4632
QY 458 AATAGATAGTTGA 470
Db 4631 ACTTCATATCTAA 4619

RESULT 11
AX347025/c
LOCUS AX347025 10329 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2096 from Patent WO0200928.
ACCESSION AX347025
VERSION AX347025.1 GI:18494911
KEYWORDS .
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A.; Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2096 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
1..10329
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 11.4%; Score 57; DB 6; Length 10329;
Best Local Similarity 48.9%; Pred. No. 0.29;
Matches 153; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 158 ATATAAATACAGTCTGATCACAACAATAATTTGATGTTGAAAAATCTACTGATTAAAC 217
Db 8678 ACAAAAATACAAACCAATAAAAAACAAAAACCTTCAAAAAATAATACCATATCTACA 8619
QY 218 TTTATAGAAAACCCAAATTTATAGCGCAACTTTTATAAACAATGTCNAACTTTCGAAGTTA 277
Db 8618 ACTATCTAATCTTTTAAAAAACCTTAAACAAAAACAAAAATAAAAAAATTCCTATATTA 8559

96619 ATTATATATTTTATTATTTTTTTTTTATTATTATTTTTTTTADATAGATATAAATAAA 96560

```

/chromosomes="12"
/clone="PFYAC357"
/clone="3D7"

```

ORIGIN

Query Match 11.1%; Score 55.6; DB 2; Length 205429;
Best Local Similarity 45.3%; Pred. No. 0.28;
Matches 202; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

Qy 16 ATTTTATAGAGAGTTAATAATAGTATATATATATATAGTAGGTACCTAATCTCTG 75
Db 102099 ATAGATATATTTTATCGATACGTAATAAATAAATAATATATTTTATATATATATA 102040

Qy 76 TCGTTAATCTAATTTAGTGTACAGAACCGACTCGTTGGCTAAGGTCTCTCCTTTTGAA 135
Db 102039 TATAAATTTTACTTTTGAATATATTTAATGAAATTTTAAATAATGCGATACAAA 101980

Qy 136 ACGCAAGCTTCTACTTTTCATGTATATAAATACAGCTGTATCACAACAACAATTTGATGA 195
Db 101979 ATGAACATTTATACTATTTTATATATATATTTAATAAATAAATGTATACATAATATA 101920

Qy 196 TTGAAATATCTACTGTANTTTAACTTTTATAGAAACCCAAATTTATAGAGCGACCTTTATA 255
Db 101919 TATATATATATATATCTTTTAAATATTTAATATATATAAATATATATATATATATAT 101860

Qy 256 AACATGTCAAACCTTCGAAGTTAAATTTAAGACCCCATATTTTACAAATTTATAGATTTTA 315
Db 101859 AATATATATATATATATATATATATTTAAGCCCTGGTACTTACATAAATATATATATA 101800

Qy 316 ATACTCCAACATTTTGTGTATGTTAAAGAGTATCCGAGTCTTTTCTTTCCAGTTTCCC 375
Db 101799 TTATATATAATTTTATATATTTTAAATAAAGTAAATTAATTTAAATAATTTATCTCA 101740

Qy 376 CACGTCCTCCAGTCTCCCGCCAGTAGAATAAGCCCAAAAGTAAACAATAAGTCGTT 435
Db 101739 CAAGTTTTTTTCTAAATATATATCCATTTAGAAAAGATTACATCTTTCCATATATTGTA 101680

Qy 436 AAAAAAGTTAAATTAATAAATAAATA 461
Db 101679 AATATCTAGTAATGAATTAATA 101654

Search completed: July 12, 2005, 05:59:32
Job time : 2541.52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 382.755 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929C-1_COPY_9829_10329
Perfect score: 501
Sequence: 1 gggctctagttattgatttt.....gtgatttgattgaacaatt 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	501	100.0	10329	4	AAA89385 Arabidops
2	501	100.0	10329	4	AAA89353 Arabidops
3	224.4	44.8	6571	4	AAA89354 Arabidops
C 4	57.8	11.5	5690	4	AAS45368 Chemicall
C 5	57.8	11.5	5690	6	Abk28205 DNA trans
C 6	57	11.4	5153	6	ABL32955 Human imm
C 7	57	11.4	10329	6	ABL34123 Human imm
C 8	53.6	10.7	18512	6	ABL32977 Human imm
C 9	51.2	10.2	392	5	ADL43653 Human ova
C 10	51.2	10.2	13427	6	ABL33927 Human imm
C 11	50.8	10.1	7467	13	ADS89706 Oligonuc
C 12	50.6	10.1	6591	6	ABL33250 Human imm
C 13	50	10.0	393	8	ABX39417 Bovine ES
C 14	50	10.0	7131	6	Abk31451 Signal tr
C 15	50	10.0	7131	6	ABL70428 Chemicall
C 16	50	10.0	7131	6	AAS61361 Human gen
C 17	50	10.0	7455	6	ABL33758 Human imm
C 18	49.6	9.9	5739	6	ABL32718 Human imm
C 19	49.6	9.9	47108	6	Abk31511 Signal tr
C 20	49.4	9.9	17137	6	ABL32190 Human imm

C 21	49.2	9.8	7457	13	ADS89432	AdS89432 Oligonuc
C 22	49.2	9.8	7534	6	ABN80154	Abn80154 Human che
C 23	49.2	9.8	10467	6	ABK28454	Abk28454 DNA trans
C 24	49.2	9.8	10710	6	ABL32892	AbL32892 Human imm
C 25	49	9.8	6014	6	AAS61404	Aas61404 Human gen
C 26	49	9.8	19380	6	AAS61426	Aas61426 Human gen
C 27	48.8	9.7	9415	6	ABL32294	AbL32294 Human imm
C 28	48.8	9.7	15743	6	ABK28396	Abk28396 DNA trans
C 29	48.6	9.7	5945	6	ABL32084	AbL32084 Human imm
C 30	48.6	9.7	6132	6	ABL32863	AbL32863 Human imm
C 31	48.6	9.7	6731	6	ABK39963	Abk39963 Human che
C 32	48.6	9.7	10891	6	ABL32464	AbL32464 Human imm
C 33	48.4	9.7	5576	6	ABL33517	AbL33517 Human imm
C 34	48.4	9.7	5576	6	ABK28262	Abk28262 DNA trans
C 35	48.4	9.7	6464	6	ABL32514	AbL32514 Human imm
C 36	48.4	9.7	9095	6	ABQ67061	Abq67061 Human ang
C 37	48.4	9.7	13420	6	ABL32917	AbL32917 Human imm
C 38	48.2	9.6	561	13	ACN46759	Acn46759 Cotton pr
C 39	48.2	9.6	6983	6	ABL32523	AbL32523 Human imm
C 40	48	9.6	7459	6	ABK31383	Abk31383 Signal tr
C 41	47.8	9.5	420	5	ADL38882	AdL38882 Human ova
C 42	47.8	9.5	421	5	ADI72427	Adi72427 Human ova
C 43	47.4	9.5	5979	4	AAS45313	Aas45313 Chemicall
C 44	47.4	9.5	5979	6	ABK28152	Abk28152 DNA trans
C 45	47.4	9.5	5987	10	ADB54308	AdB54308 Pretreat

ALIGNMENTS

RESULT 1
AAA89385
ID AAA89385 standard; DNA; 10329 BP.

AC AAA89385;

DT 04-NOV-2004 (revised)

DT 23-APR-2001 (first entry)

XX Arabidopsis thaliana gene involved in epigenetic gene silencing.

XX Gene silencing; silencing gene; MOM; ds.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT exon

FT /tag= a

FT /number= 1

FT /tag= b

FT /number= 1

FT /tag= c

FT /number= 2

FT /tag= d

FT /product= "Columbia protein"

FT /note= "this sequence contains 15 introns."

FT /tag= e

FT /number= 2

FT /tag= f

FT /number= 3

FT /tag= g

FT /number= 3

FT /tag= h

FT /number= 4

FT /tag= i

FT /number= 1

FT /number= 4
FT 3507. .3680
FT /*tag= j
FT /number= 5
FT intron 3681. .3973
FT /*tag= k
FT /number= 5
FT exon 3974. .4895
FT /*tag= l
FT /number= 6
FT allele 4338
FT /*tag= m
FT /note= "ecotype Zurich"
FT intron 4896. .4975
FT /*tag= n
FT exon 4976. .5217
FT /*tag= o
FT /number= 7
FT intron 5218. .5777
FT /*tag= p
FT /number= 7
FT exon 5778. .5882
FT /*tag= q
FT /number= 8
FT intron 5883. .6082
FT /*tag= r
FT /number= 8
FT exon 6083. .7480
FT /*tag= s
FT /number= 9
FT allele 6721
FT /*tag= t
FT /note= "ecotype Zurich"
FT intron 7481. .7615
FT /*tag= u
FT /number= 9
FT exon 7616. .7771
FT /*tag= v
FT /number= 10
FT intron 7772. .7914
FT /*tag= w
FT /number= 10
FT exon 7915. .8070
FT /*tag= x
FT /number= 11
FT intron 8071. .8153
FT /*tag= y
FT /number= 11
FT exon 8154. .8318
FT /*tag= z
FT /number= 12
FT intron 8319. .8451
FT /*tag= aa
FT /number= 12
FT exon 8452. .8629
FT /*tag= ab
FT /number= 13
FT intron 8630. .8718
FT /*tag= ac
FT /number= 13
FT exon 8719. .8918
FT /*tag= ad
FT /number= 14
FT intron 8919. .9000
FT /*tag= ae
FT /number= 14
FT exon 9001. .9211
FT /*tag= af
FT /number= 15
FT intron 9212. .9284
FT /*tag= ag
FT /number= 15

FT exon 9285. .10329
FT /*tag= ah
FT /number= 16
FT intron WO200100801-A2.
FT exon 04-JAN-2001.
FT intron 21-JUN-2000; 2000WO-EP005761.
FT exon 23-JUN-1999; 99GB-00014623.
FT intron (NOVS) NOVARTIS AG.
FT exon (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT intron Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
FT exon WPI; 2001-137952/14.
FT intron P-PSDB; AAB20063.
FT exon Novel gene encoding a protein that controls gene silencing, in particular
FT intron silencing of plant genes.
FT exon Claim 5; Page; 49pp; English.
FT intron
FT exon The present sequence is that of an Arabidopsis thaliana ecotype Columbia
FT intron gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino
FT exon acid protein with similarity with ATPase/helicase proteins of the
FT intron SWI2/SNF2 family, and which controls gene silencing, particularly plant
FT exon gene silencing. Trans-acting modified loci were identified by T-DNA
FT intron insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
FT exon carries a heritable inactivated, methylated hygromycin resistance gene
FT intron (hyg). Mutation of a silencing modified locus results in release of
FT exon silencing of the hyg gene and restores hygromycin resistance. Plants
FT intron homozygous for the silent resistance gene were subjected to
FT exon transformation with a bar selectable marker gene under control of the T-
FT intron DNA 1'-2' dual promoter. Transformants were selected and their progeny
FT exon screened for hygromycin resistance. The mutant phenotype (hygromycin
FT intron resistance) was screened for genetic co-segregation with a specific T-DNA
FT exon insert. Cloning of the tagged gene allowed characterization of the
FT intron present sequence. 2 Allelic differences were found in the corresponding
FT exon gene of A. thaliana ecotype Zurich, occurring at nucleotide positions
FT intron 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful
FT exon as a molecular tool for regulating gene expression. Note: the present
FT intron sequence is not shown in the specification but is derived from the A.
FT exon thaliana mutant silencer gene sequence given in AAA89353
FT intron
FT exon Revised record issued on 04-NOV-2004 : Correction to feature table key
FT intron
FT exon Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;
FT intron
FT exon Query Match 100.0%; Score 501; DB 4; Length 10329;
FT intron Best Local Similarity 100.0%; Pred. No. 1.1e-93;
FT exon Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT intron
FT exon 1 GGGTCTTAGTTATTCGATTTTATAGAGAGTTAAATAGTAGTATATATATATATAGTAGTAG 60
FT intron |||||
FT exon 9829 GGGTCTTAGTTATTCGATTTTATAGAGAGTTAAATAGTAGTATATATATATAGTAGTAG 9888
FT intron |||||
FT exon 61 GTTACCTAATCTCTGCTTAATCTAATTTAGTAGTACGAGAACCGACTCGTTGGCTAAG 120
FT intron |||||
FT exon 9889 GTTACCTAATCTCTGCTTAATCTAATTTAGTAGTACGAGAACCGACTCGTTGGCTAAG 9948
FT intron |||||
FT exon 121 GTCTCTCCTTTTGAACCGCAACGGTCTTACTTTTCATGTATATAAATACAGTCTGATCACAC 180
FT intron |||||
FT exon 9949 GTCTCTCCTTTTGAACCGCAACGGTCTTACTTTTCATGTATATAAATACAGTCTGATCACAC 10008
FT intron |||||
FT exon 181 AACACAAATTGATGATTTGAAATATCTACGATTTAACTTTATAGAAAACCCAAATTTATAG 240
FT intron |||||
FT exon 10009 AACACAAATTGATGATTTGAAATATCTACTGATTTAACTTTATAGAAAACCCAAATTTATAG 10068
FT intron |||||
FT exon 241 AGCGACAACCTTTTATAAACAATGTCAAACTTCGAAGTTAAAATTTAAAGACCCCAATTTTA 300
FT intron |||||

XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
DR WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 6; Page 19-22; 48pp; English.
XX
CC The present sequence is that of an Arabidopsis thaliana gene (MOM)
CC involved in epigenetic gene silencing. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia, occurring at nucleotide positions
CC 4338 (A instead of T) and 6721 (T instead of G). Gene silencing is useful
CC as a molecular tool for regulating gene expression
CC
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
XX Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 501; DB 4; Length 10329;
Best Local Similarity 100.0%; Pred. No. 1.1e-93;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGTCTTAGTTATGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATATAGTAG 60
DB 9829 GGGTCTTAGTTATGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATATAGTAG 9888
QY 61 GTTACCTAATCTGCTGTTAACTAATTTAGTGAGTCAGAACCGACTCGTTGGCTAAG 120
DB 9889 GTTACCTAATCTGCTGTTAACTAATTTAGTGAGTCAGAACCGACTCGTTGGCTAAG 9948
QY 121 GTCTCTCTTTTGAACCGCAACGTTCTACTTTTCATGTATATAATACAGTCTGATCACAC 180
DB 9949 GTCTCTCTTTTGAACCGCAACGTTCTACTTTTCATGTATATAATACAGTCTGATCACAC 10008
QY 181 AACACAAATGATGATGAAATACTACTGATTTAATGATGATGATGATGATGATGATGATGAT 240
DB 10009 AACACAAATGATGATGAAATACTACTGATTTAATGATGATGATGATGATGATGATGATGAT 10068
QY 241 AGCGACAACTTTAATACATGTCACAACTTCGAGTTAAATTTTAAAGACCCCAATATTTTA 300
DB 10069 AGCGACAACTTTAATACATGTCACAACTTCGAGTTAAATTTTAAAGACCCCAATATTTTA 10128
QY 301 CAATTATAGATTTTAAATACCTCCAACTATTTTGTGATGTTAAAGAAAGTATCCGAGTCTTT 360
DB 10129 CAATTATAGATTTTAAATACCTCCAACTATTTTGTGATGTTAAAGAAAGTATCCGAGTCTTT 10188
QY 361 TCTTTCCAGTTTCCCAACCGTCCCATGATCTCCCGCCAGTAGTAAAGAAAGTATCCGAGTCTTT 420
DB 10189 TCTTTCCAGTTTCCCAACCGTCCCATGATCTCCCGCCAGTAGTAAAGAAAGTATCCGAGTCTTT 10248
QY 421 AAACAAAGAGTCTGTTAAAGAAAGTAAATTTAAAGAAAGTAAATTTAAAGAAAGTAAATTTAA 480
DB 10249 AAACAAAGAGTCTGTTAAAGAAAGTAAATTTAAAGAAAGTAAATTTAAAGAAAGTAAATTTAA 10308

QY 481 AGTGATTTGAATTTGAACAATT 501
DB 10309 AGTGATTTGAATTTGAACAATT 10329
RESULT 3
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
XX
AC AAA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT 310..6315
FT CDS /*tag= a
XX
XX WO200100801-A2.
XX
XX 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-BP005761.
XX
XX 23-JUN-1999; 99GB-00014623.
XX
XX (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
XX P-PSDB; AAB20062.
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 6; Page 22-32; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression
XX
SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;
Query Match 44.8%; Score 224.4; DB 4; Length 6571;
Best Local Similarity 97.4%; Pred. No. 8.2e-37;
Matches 228; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GGGTCTTAGTTATGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATAGTAG 60
DB 6336 GGGTCTTAGTTATGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATAGTAG 6395

[illegible]

RESULT 4	
AAS45368/c	
ID	AAS45368 standard; DNA; 5690 BP.
XX	
XX	
AC	AAS45368;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Chemically pretreated genomic DNA associated with cell cycle #37.
XX	
KW	Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW	graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW	immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW	PCR primer.

SQ	Sequence	5690 BP; 1509 A; 124 C; 1264 G; 2753 T; 0 U; 40 Other;
Query Match	11.5%; Score 57.8; DB 4; Length 5690;	
Best Local Similarity	49.2%; Pred. No. 0.015;	
Matches	146; Conservative 0; Mismatches 151; Indels 0; Gaps 0;	
QY	162 AAATACAGTCTGATCACAACAACAATTTCATGATTTGAATAACTACTGATTAATTAACCTTTA	221
DB	4716 AACGAAACGGGACCCCACTAAAAATAATAATAAAAATACAAATAACAATAACGAAACTA	4657
QY	222 TAGAAAAACCMAATTATAGAGCGCACAACTTTATAAACATGTCCAACCTTCGAACTTAAAT	281
DB	4556 ACGAAAAAATATTAACGAAAAACCTTAANTANTTCAAACATTCACATTCGACTATNAAT	4597
QY	282 TTAGAGCCCCCATAAATTTTACAATTTATAGATTTTTAAATACTCCAACACTATTTTGATGTTAA	341
DB	4596 AAAAAATCCCTCACTCTAANACTTAATATACAAAATCTTTTATAATTAACACCTTTCAAT	4537
QY	342 AGAAGTATCGGAGTCTTTTCTTCAGTTTCCCACCGTCCCATGACTCCCCCAGCCAG	401
DB	4536 TAAAAAAAACATTTTTCCTCTTCATCTTCCTACCCCTACCCATATATAAATTCGCCGATAAAA	4477
QY	402 TAGAAAAAGCCAAAAAGTAACAACAAAAGTCGTTTAAAAAGTTTAAATTTAAAAAAA	458
DB	4476 CAAAAAAATACTCAAAACTAAACTCAAAATAAAAAACGAAAAATTAATAATAAAAAATAA	4420
RESULT 5		
ABK28205/C		
ID	ABK28205 standard; DNA; 5690 BP.	
XX AC	ABK28205;	
XX XX	23-APR-2002 (first entry)	
XX DT	DNA transcription associated genomic DNA #40.	
XX DE	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;	
XX KW	PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;	
XX KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;	
XX KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;	
XX KW	immunological disorder; Werner syndrome; developmental disorder;	
XX KW	psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;	
XX KW	neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;	
XX KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;	
XX KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;	
XX KW	polyglutamine disorder; solid tumour.	
XX OS	Unidentified.	
XX XX	WO200192565-A2.	
XX PN	06-DEC-2001.	
XX PD	06-APR-2001; 2001WO-EP003973.	
XX PF	06-APR-2000; 2000DE-01019058.	
XX PR	07-APR-2000; 2000DE-01019173.	
XX FR	30-JUN-2000; 2000DE-01032529.	
XX PR	01-SEP-2000; 2000DE-01043826.	
XX XX	(EPIG-) EPIGENOMICS AG.	
XX PA	Olek A, Piepenbrock C, Berlin K;	
XX PI	WPI; 2002-090046/12.	
XX DR	New nucleic acids or oligomers, useful for diagnosing or treating	
XX PT	diseases associated with DNA transcription, e.g. immunological disorders,	
XX PT	Werner syndrome, psoriasis, myocardial infarction, solid tumors or	
XX PT	cancer.	
XX PS	Claim 1; SEQ ID NO 79; 32pp; English.	

Qy 429 AGTCGTTAAAAAGTTAAATTAAAAAAAAT 460
| ||||| || ||||| |||||
Db 12964 ATATATAAAAAACTTCAATATAAAAACAAT 12913

RESULT 9
ADL43653/c
ID ADL43653 standard: DNA: 392 BP.

DT 20-MAY-2004 (first entry)

Human ovarian cancer DNA marker #17543.

Human: ovarian cancer: ds: tumour: cytostatic: DNA marker.

xx
os
Homo sapiens.

AA
PN
WO200170979-A2

27-SEP-2001

XX
PF 21-MAR-2007. 2001WO-IIS009126XX
PR 21-MAR-2000: 2000US-0191031P

PR 25-MAY-2000; 2000US-0207124P.

PR 15-JUN-2000; 2000US-0211940P.

PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P

21-DEC-2000; 2000US-0257672P.
PR

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lee J, Lillie J;

WPI; 2001-611502/70.

Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

PS Disclosure: SEO ID NO 17543: 106pp: English.

The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of a marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polynucleotide or antibody of the invention is used to treat ovarian cancer.

KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antiparasitic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP007537.
XX PF
XX 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX Claim 1; SEQ ID NO 1223; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6591 BP; 1983 A; 40 C; 1241 G; 3327 T; 0 U; 0 Other;
Query Match 10.1%; Score 50.6; DB 6; Length 6591;
Best Local Similarity 46.1%; Pred. No. 0.46;
Matches 206; Conservative 0; Mismatches 239; Indels 2; Gaps 1;
QY 54 TAAGTAGGTACCTAATCTCTGCTGTTAACTCTAATTTAGTGAGTCAGGACCGTCGTT 113
DB 2843 TAAACATTTCTCTTTATCGCTTAATTTTATATAAACTAACTTTCTTTAAACTATAT 2784
QY 114 GGCTAAGGTCTCTCTTTTGAAACGGCAACGTTCTCTATCTATATATAATACAGTCGTG 173
DB 2783 CTCTTCGATCCCTCACTACTCTTCTCTATCCATCTCTCCATCTATTTCTTAATCTTC 2724
QY 174 ATCACACACACAAATGATGATGAAATACATCTAGTAACTTTATAGAAAACCCAA 233
DB 2723 TCTACAAATCCCATTTATATATAATATAATAAAATTTATTAACATATCTAAAAATCCAA 2664
QY 234 ATTATAGAGGACAACTTTATAACATGTCAAACTTCGAGTTAAATTTAAGACCCCAT 293
DB 2663 ACTCTTTCTCTCTCTATACCTTTACTCCCATTTTCAAAACAACTTTATAATATACC 2604
QY 294 AATTTTACAATTTAGATTTTATATCTCCCACTATTTTGTGATGTTTAAAGAGAGTATCCG 353
DB 2603 AA-CTTCAATCATATATTTTATATAAACTTAACCTAAACGATTAACCAAAATAA 2546
QY 354 AGTCTTTCTTTCCAGTTTCCCAACCGTCCCATGACTCCCGCCAGTCAGTAAAGACCA 413
DB 2545 AAACTTTACATTAATAAAACCAAAATTAATCTATCTAAAAAACAACAAAAAATAAT 2486
QY 414 AAAAGTAAACAAAGTCGTTAAAAAGTTAAATTAATAAAATAAATAAATAGATAGTTCAGT 473
DB 2485 AAAAAATAAAAAAATAAAAAACGAAAAAATAAATAAATAAATAAATAAATAAATAA 2426

QY 474 TTACTAAAGTGAATTTGAATTTGAACAAAT 500
DB 2425 TTACTAAATAAATTTTAAATAAAAAACT 2399

RESULT 13

ABX39417/c
ID ABX39417 standard; cDNA; 393 BP.
XX AC ABX39417;
XX DT 20-FEB-2003 (first entry)
XX DE Bovine EST associated with lactation/muscle/fat deposition #4582.
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX OS Bos Taurus.
XX US2002137139-A1.
XX PD 26-SEP-2002.
XX PF 24-SEP-2001; 2001US-00960352.
XX PR 12-JAN-1999; 99US-0115707P.
XX PR 11-JAN-2000; 2000US-00480902.
XX (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
PI WPI; 2003-110599/10.
XX New nucleic acid associated with lactation, and muscle and fat
XX deposition, useful for genome mapping, gene identification and analysis,
XX cattle breeding, or for genetically improving cattle.
XX Claim 2; SEQ ID NO 4582; 245pp; English.
XX The invention relates to a purified nucleic acid molecule associated with
XX lactation or muscle and fat deposition (designated LMFD), derived from
XX cattle, and the LMFD nucleic acid can specifically hybridise to a second
XX nucleic acid molecule comprising any of 15112 nucleotide sequences,
XX appearing as ABX34836-ABX49947, or complements of them. Also included are
XX : (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
XX acid linked to a promoter and a 3' non-translated sequence that
XX functions in the cell to cause termination of transcription and addition
XX of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
XX (2) determining a level or pattern of a molecule in a bovine cell or
XX tissue comprising: (a) incubating a marker nucleic acid (comprising any
XX of the 15112 nucleic acid sequences or its complement or fragment) with a
XX complementary nucleic acid molecule obtained from the bovine cell or
XX tissue, where hybridisation between the marker nucleic acid and the
XX complementary nucleic acid permits the detection of the molecule; and (b)
XX detecting the level or pattern of the complementary nucleic acid, where
XX the detection of the complementary nucleic acid is predictive of the
XX level or pattern of the molecule. The LMFD nucleic acid is used for
XX determining a level or pattern of a molecule in a bovine cell or tissue.
XX It is useful for genome mapping, gene identification and analysis, cattle
XX breeding, preparation of constructs for use in cattle gene expression, or
XX for genetically improving cattle. The present sequence is one of the
XX 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
XX present sequence was not shown in the specification but was obtained in
XX electronic format from the USPTO web site:
XX seqdata.uspto.gov/sequence.html?DocID=20020137139

[illegible]

Search completed: July 12, 2005, 01:45:57
Job time : 385.755 secs

C	1	47.6	9.5	832	4	US-09-621-976-2813	Sequence 2813, A
C	2	47.4	9.5	601	4	US-09-949-016-86304	Sequence 86304, A
C	3	47.4	9.5	1137	4	US-09-248-796A-3501	Sequence 3501, Ap
C	4	46.6	9.3	298336	4	US-09-949-016-15600	Sequence 16600, A
C	5	46.2	9.2	601	4	US-09-949-016-86305	Sequence 86305, A
C	6	46.2	9.2	422992	4	US-09-949-016-14182	Sequence 14182, A
C	7	44.4	8.9	3255	4	US-09-949-016-14182	Sequence 3524, Ap
C	8	43.6	8.7	126176	4	US-09-949-016-16137	Sequence 16137, A
C	9	43.6	8.7	126176	4	US-09-949-016-16138	Sequence 16138, A
C	10	42.8	8.5	6124	3	US-08-213-415B-3	Sequence 3, Appli
C	11	42.6	8.5	2435	3	US-09-308-593-1	Sequence 1, Appli
C	12	42.4	8.5	1039	4	US-09-902-540-1280	Sequence 1280, Ap
C	13	41.8	8.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
C	14	41.4	8.3	99830	4	US-09-949-016-16859	Sequence 16859, A
C	15	41.2	8.2	136917	4	US-09-949-016-16369	Sequence 16369, A
C	16	41	8.2	538	4	US-09-621-976-9313	Sequence 9313, Ap
C	17	41	8.2	37950	3	US-09-338-907-183	Sequence 183, App
C	18	41	8.2	37950	3	US-09-218-207-183	Sequence 183, App
C	19	40.8	8.1	1141	4	US-09-806-709B-22	Sequence 22, Appli
C	20	40.8	8.1	4285	3	US-09-410-464-1	Sequence 1, Appli
C	21	40.8	8.1	44353	4	US-09-949-016-15302	Sequence 15302, A
C	22	40.6	8.1	38575	4	US-09-949-016-12876	Sequence 12876, A
C	23	40.6	8.1	38575	4	US-09-949-016-13871	Sequence 13871, A
C	24	40.6	8.1	134987	4	US-09-949-016-13348	Sequence 13348, A
C	25	40.6	8.1	134987	4	US-09-949-016-15349	Sequence 15349, A
C	26	40.6	8.1	134987	4	US-09-949-016-15350	Sequence 15350, A
C	27	40.6	8.1	134987	4	US-09-949-016-15507	Sequence 15507, A

C	1	47.6	9.5	832	4	US-09-621-976-2813	Sequence 2813, A
C	2	47.4	9.5	601	4	US-09-949-016-86304	Sequence 86304, A
C	3	47.4	9.5	1137	4	US-09-248-796A-3501	Sequence 3501, Ap
C	4	46.6	9.3	298336	4	US-09-949-016-15600	Sequence 16600, A
C	5	46.2	9.2	601	4	US-09-949-016-86305	Sequence 86305, A
C	6	46.2	9.2	422992	4	US-09-949-016-14182	Sequence 14182, A
C	7	44.4	8.9	3255	4	US-09-949-016-14182	Sequence 3524, Ap
C	8	43.6	8.7	126176	4	US-09-949-016-16137	Sequence 16137, A
C	9	43.6	8.7	126176	4	US-09-949-016-16138	Sequence 16138, A
C	10	42.8	8.5	6124	3	US-08-213-415B-3	Sequence 3, Appli
C	11	42.6	8.5	2435	3	US-09-308-593-1	Sequence 1, Appli
C	12	42.4	8.5	1039	4	US-09-902-540-1280	Sequence 1280, Ap
C	13	41.8	8.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
C	14	41.4	8.3	99830	4	US-09-949-016-16859	Sequence 16859, A
C	15	41.2	8.2	136917	4	US-09-949-016-16369	Sequence 16369, A
C	16	41	8.2	538	4	US-09-621-976-9313	Sequence 9313, Ap
C	17	41	8.2	37950	3	US-09-338-907-183	Sequence 183, App
C	18	41	8.2	37950	3	US-09-218-207-183	Sequence 183, App
C	19	40.8	8.1	1141	4	US-09-806-709B-22	Sequence 22, Appli
C	20	40.8	8.1	4285	3	US-09-410-464-1	Sequence 1, Appli
C	21	40.8	8.1	44353	4	US-09-949-016-15302	Sequence 15302, A
C	22	40.6	8.1	38575	4	US-09-949-016-12876	Sequence 12876, A
C	23	40.6	8.1	38575	4	US-09-949-016-13871	Sequence 13871, A
C	24	40.6	8.1	134987	4	US-09-949-016-13348	Sequence 13348, A
C	25	40.6	8.1	134987	4	US-09-949-016-15349	Sequence 15349, A
C	26	40.6	8.1	134987	4	US-09-949-016-15350	Sequence 15350, A
C	27	40.6	8.1	134987	4	US-09-949-016-15507	Sequence 15507, A

RESULT 2
US-09-949-016-86304/c
; Sequence 86304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86304
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-86304

Query Match 9.5%; Score 47.4; DB 4; Length 601;
Best Local Similarity 52.3%; Pred. No. 0.021; Indels 0; Gaps 0;
Matches 102; Conservative 1; Mismatches 92;
QY 148 ACTTTCATGTATATAATACAGTCTGATCACACACAAATTTGATGATTGAAATPACTA 207
DB 447 ACATTTTGTATATAAATAAATAAATAAACAATCTTAATATATAAATAAACAAT 388
QY 208 CTGATTTAACTTTATAGAAAACCCAAATTTATAGAGCGACAACTTTTAAACATGTCAAC 267
DB 387 TATATAATAATATAGAAAATTTATATATTTTAAATATATAAACAAGTATATT 328
QY 268 TTCGAGTTTAAATTTAAGACCCCAATAATTTTACAATTTATAGATTTTAACTCCAATA 327
DB 327 TTATATATTTAAATATATAAACAAGTATTTTATATATTTTAAATATATAAACAAGTATA 268
QY 328 TTTTGTGATGTAAA 342
DB 267 TTTTATATATTTAAA 253

RESULT 3
US-09-248-796A-3501
; Sequence 3501, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3501
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3501

Query Match 9.5%; Score 47.4; DB 4; Length 1137;
Best Local Similarity 51.7%; Pred. No. 0.025; Indels 0; Gaps 0;
Matches 108; Conservative 0; Mismatches 101;

QY 153 CATGTATATAATACAGTCTGATCACACACAAATTTGATGATTGAAATPACTACTGAT 212
DB 110 CAGGATTATTCAGTTTATTATATTTTGAACACAAATTAATAATGATTATTAGACACACAG 169
QY 213 TTAACCTTTATAGAAAACCCAAATTTATAGAGCGACAACTTTTAAACATGTCAAACTTCCA 272
DB 170 TCAAAAATAAAGAAAACCGACGAGAAGGGAAGTGGGAATATTAAAGAAATTTCAAGCTGCCA 229
QY 273 AGTTAAATTTTAAAGACCCCATAAATTTTACAATTTATAGATTTTAAATACTCCAACTATTTTG 332
DB 230 ATCCCTCATTTAAAGAACTTGTCTTTAATCATATATGAAAGSCATTTATTGAATTCATTAG 289
QY 333 TGATGTTTAAAGAAAGTATCCGAGTCTTTT 361
DB 290 TGACACCCAGAGAAGATAGCAGTTACATTT 318

RESULT 4
US-09-949-016-16600/c
; Sequence 16600, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16600
; LENGTH: 298336
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(298336)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16600

Query Match 9.3%; Score 46.6; DB 4; Length 298336;
Best Local Similarity 45.5%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 166; Conservative 0; Mismatches 199;
QY 4 TCTTATGTTATTGATTTTATAGAGATTAAATAGTATATATATATATATATATATATAGTAGGTT 63
DB 274210 TTTTATTTTATATATTTTATATATATATATATATATATATATATATATATATATATTT 274151
QY 64 ACCTAATCTCTGCTGTTAATCTAATTTAGTGTAGGAAACCGACTGCTGGTGAAGGTC 123
DB 274150 ATATAATATATAAATTTATGTA 274091
QY 124 TCTCCTTTTGAACCGCAACGTTCTACTTTTCAGTATATAAATACAGTCTGTACACACAAC 183
DB 274090 TATATTATATATATGTTATATATAAATATATATATTTCTATATTTTATATAAATATATATATT 274031
QY 184 ACAAAATTTGATTTGAAAATACCTACTGATTTAACTTTTATAGAAAACCCAAATTTATAGAGC 243
DB 274030 ATATATTTATATATAAATAATATATATTTCTATAAATAATAATAATAATAATAATAATA 273971
QY 244 GACAACTTTATAAACATGCAAACTTCGAAAGTTAAATTTAAGACCCCAATATTTTACAA 303
DB 273970 TTATTATATAAATAATATGTTTATATACGTGTATATATATATATATATATATATATATATAG 273911
QY 304 TTATAGATTTTAACTACTCCAACTATTTTGTGATGTTTAAAGAAAGTATCCGAGTCTTTTCT 363

Db 273910 TTATATATTATACGTTGATATATATATATATACAGTATAGATATATAGTTATATATAT 273851
Qy 364 TTCCA 368
Db 273850 ATACA 273846

RESULT 5

US-09-949-016-86305/c
; Sequence 86305, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86305
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-86305

Query Match 9.2%; Score 46.2; DB 4; Length 601;
Best Local Similarity 52.3%; Pred. No. 0.041;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Qy 148 ACTTTCATGTATATAATACAGTCTGTATACACACACAAATTCATGATGATGAAATACTA 207
Db 478 ACATTTTGTATATTAATAATAATAATAAAACATATCTTAATATATTAATAATAAAACAT 419
Qy 208 CTGATTTAACTTTATAGAAACCCAAATTTATAGCGACAACTTTATACATGTCACAAAC 267
Db 418 TATATATAATATAGAAATTTATATTTTATATTTTAAATATATAAAACAGTATATT 359
Qy 268 TTCCGAGTTTAAATTTTAAGACCCCATTAATTTTACAATTTATAGATTTTAACTCCAACTA 327
Db 358 TTATATATTAAATATATAAACAAGTCTATTTATATATTAAATATATAAAACAGTATA 299
Qy 328 TTTTGTGATGTTAAA 342
Db 298 TTTTATATATTAAA 284

RESULT 6

US-09-949-016-14182
; Sequence 14182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14182

; LENGTH: 422592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(422592)
; OTHER INFORMATION: n = A, T, C or G
US-09-949-016-14182
Query Match 9.2%; Score 46.2; DB 4; Length 422592;
Best Local Similarity 52.3%; Pred. No. 0.22;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Qy 148 ACTTTCATGTATATAATACAGTCTGTATACACACACAAATTCATGATGATGAAATACTA 207
Db 333134 ACATTTTGTATATTAATAATAATAATAAAACATATCTTAATATATTAATAATAAAACAT 333193
Qy 208 CTGATTTAACTTTATAGAAACCCAAATTTATAGCGACAACTTTATACATGTCACAAAC 267
Db 333194 TATATATAATATAGAAATTTATATTTTATATTTTAAATATATAAAACAGTATATT 333253
Qy 268 TTCCGAGTTTAAATTTTAAGACCCCATTAATTTTACAATTTATAGATTTTAACTCCAACTA 327
Db 333254 TTATATATTAAATATATAAACAAGTCTATTTTATATATTAAATATATAAAACAGTATA 333313
Qy 328 TTTTGTGATGTTAAA 342
Db 333314 TTTTATATATTAAA 333328

RESULT 7

US-09-710-279-3524
; Sequence 3524, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P0348005
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3524
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3524

Query Match 8.9%; Score 44.4; DB 4; Length 3255;
Best Local Similarity 58.2%; Pred. No. 0.18;
Matches 78; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
Qy 129 TTTTGAACGCAACGTTCTACTTTTCATGTATATAAATACAGTCTGATCACAACACAAA 188
Db 2049 TTGAGAAAGTGAGCTTTCATTTATCAATTTAAAGATTTTGAACAATTCATTAATAATA 2108
Qy 189 TTGATGATTTGAAATATCTACTCTGATTTTAACTTTATAGAAAAACCCCAATTTATAGCGCA 248
Db 2109 TTATTGTTAAATTAATAAATCTTAATAAGTGTGTTTTAAAGTTTATATTATAGGACCAAGTA 2168
Qy 249 CTTTATAAATGCT 262
Db 2169 CGTTATAATGAAGT 2182

RESULT 8

US-09-949-016-16137
; Sequence 16137, Application US/09949016

QY 260 TGTCAACTTCGAAGTTAAATTTAAGACCCCAATAATTTTACAAATTATAGATTTTAAATAC 319
Db 2326 AAGAAAAATTAATAATCAACATATTCAAAAAATTTAAAGTTCTTAAAAATATTATATA 2385
QY 320 TCCAACTATTTTGTGATGTTAAAGAGATCCGAGTCTTTTCTTCCAGTTTCCCAACC 379
Db 2386 ACTTAATATTCATATTCATTAATGAAGTCAATATTTCTTGTGTTTTTTCATATTTGTGTAAG 2445
QY 380 GTCCCATGATCTCCCGAGCAGTAGAAAAAGCCAAAGCAAAAGTAAACAAAGTCTGTTAAAA 439
Db 2446 AATGAAAAAATAAAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAA 2505
QY 440 AGTTAAATTAATAATAATAATAGATAGTTGACGTTTACTTAA 481
Db 2506 AAGAAAAACAATATGTAAAAATATAATTTATATATAATAA 2547

RESULT 11
US-09-306-593-1
; Sequence 1, Application US/09306593
; Patent No. 6184018
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; APPLICANT: Ximenes, Eduardo A.
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
; FILE REFERENCE: 31-98us
; CURRENT APPLICATION NUMBER: US/09/306,593
; CURRENT FILING DATE: 1999-05-06
; EARLIER APPLICATION NUMBER: US 60/084,494
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2435
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39)..(2009)
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: (87)..(2009)
US-09-306-593-1

Query Match 8.5%; Score 42.6; DB 3; Length 2435;
Best Local Similarity 47.7%; Pred. No. 0.45;
Matches 156; Conservative 0; Mismatches 169; Indels 2; Gaps 1;
QY 140 AAGTTCTACTTTTCATGTATATAAATACAGTCTGATCAGACACACACAAATTTGATGTTGA 199
Db 1969 AACCTCTGGTCAATGGTACCTCGAAATGTTGACAAACAACTAAATTTCTTAAAAATTTA 2028
QY 200 AAATACTACTGATTTAACTTTATAGAAAAACCCAAATTTATAGACGCGACAACTTTATAAACA 259
Db 2029 TAATAATATTTTATTACATTAATAATAATAATATATAATTAATGGAATTTTATTCACT 2088
QY 260 TGTCAAAC--TTCGAAGTTAAATTTAAGACCCCAATAATTTTACAAATTATAGATTTTAAAT 317
Db 2089 TCTTTTGTCTATAAGTAGTAGTGAATAAATTAATTTTATAAATATATATAAATTTATAGAATAA 2148
QY 318 ACTCCAACATTTTGTGTTTAAAGAGATCCGAGTCTTTCTTCCAGTTTCCCA 377
Db 2149 TCTTTTGTGAATCATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATA 2208
QY 378 CCGTCCCATGACTCCCGCCAGTAGAGAAAAAGCCAAAAAGTAAAAAGTAAAGTCCGTAA 437
Db 2209 TTATTTAAATATTTCTTAAGAGAGATTTATATTTTATAGNATTAATATAAAGCAAGAA 2268
QY 438 AAAAGTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 464

Db 2269 ACAATATATATTTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATA 2295

RESULT 12
US-09-902-540-1280
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 8.5%; Score 42.4; DB 4; Length 1039;
Best Local Similarity 45.5%; Pred. No. 0.41;
Matches 148; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 157 TATATAATACAGTCTGATCAGACACACACAAATTTGATGATGATGATGATGATGATGATGAT 216
Db 240 TATATA 239
QY 217 CTTTATAGAAAAACCCAAATTTATAGACGCGACAACTTTTATAACATGTCAAACTTCGAAGTT 276
Db 300 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 359
QY 277 AAAATTTAAGACCCCAATAATTTTACAAATTATAGATTTTATTAATCTCAACTATTTTGTGAT 336
Db 360 ATAACAAAAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 419
QY 337 GTTAAAAAGAGTATCCGAGTCTTTTCTTCTCCAGTTTCCCGCCGCTCCCATGACTCCCA 396
Db 420 NTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 479
QY 397 GCCAGTAGAAAAAGCCAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 456
Db 480 AATA 539
QY 457 AAATAGATAGTTGACGTTTACTTAA 481
Db 540 AAACAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 564

RESULT 13
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm

```

; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CD
; LOCATION: 23
; US-09-621-976 28

```

Query Match	8.3%;	Score 41.8;	DB 4;	Length 832;	
Best Local Similarity	10.1%;	Pred. No. 0.54;			
Matches 37;	Conservative 168;	Mismatches 160;	Indels 0;	Gaps 0;	
Qy	77	CGTTAAATCTAAATTTAGTGAGTCAGAAACCGAGCTCGTTGGCTAAAGTCTCTCTTTTTCGAAA	136		
Db	1	YRWYWKYITWYAKWTGKWKWSWSYMWYKWKYWRWRKKKKAWKWKYKTWTWYWW	60		
Qy	137	CGCAACCTGTCATCTTCCTGATPATATAAATACAGTCTGATCATCACACACAAATGATGAT	196		
Db	61	RYAMWGTYKKKAMCRCTTKKKKKGGYMMWYMGWRSSYMAWMTRTWTGAYAYRSMYWW	120		
Qy	197	TGAAAACTACTGATTAACTTTATAGAAAACCAAAATTTATAGAGCGACAACTTTATAA	256		
Db	121	YRCWKKAYIRKTYCTYSKGTWWRKWRKGAWTTWWKKYITWATRTYWMCMCTKRRWAS	180		
Qy	257	ACATGTCAAACTTCGAAGTTAAAAATTTAAGACCCCATAAATTTTACAATTTATAGATTTTAA	316		
Db	181	WYCWNGWKARKWSTWRKRSYASARSASRKCYSCSGMSGMSWKYMRWRMGWATGAGM	240		
Qy	317	TACTCCAACTATTTTGTGATGTTTAAAGAAGATATCCGAGTCTTTTCTTTCAGTTTCCCC	376		
Db	241	KAWRASCMRRKYAGKSKTSYKSMWCMWTRSMWKCYTKARWTGYCYRKGGMGWKGRWY	300		
Qy	377	ACCGTCCCATGACTCCCCCAGCAGTAGAAAAAGCCAAAAAGTAAACAAAAAGTCGTTA	436		
Db	301	ASKYMKRKKWVWCWARMYRISTGTGRASWWRWRYTMMKKWKYAWARAAWVWAWWAW	360		
Qy	437	AAAAA 441			
Db	361	RRACA 365			

```

RESULT 14
US-09-949-016-16859
; Sequence 16859, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16859
; LENGTH: 99830
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(99830)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16859

```

Query Match 8.3%; Score 41.4; DB 4; Length 99830;
Best Local Similarity 48.9%; Pred. No. 2.3;

	Matches	111;	Conservative	0;	Mismatches	116;	Indels	0;	Gaps	0;
Qy	33	AATAGTATATATATATATGTAATACTACGGTTACTTAATCTCTGTCGTAAATCTCAATTTAG	92							
Db	95236	AATAGTGTTATATGTCAGGTTATATATAAATAAACACATATAAGTTTTCATGTTATCTTT	95295							
Qy	93	TGAGTCAGAAACCGACTCGTTGGCTTAAGGTCTCTCCTTTTTGAAGCGCAACGTTCTCACTTT	152							
Db	95296	GGGTGCAGACAGCTTTTCTGTGGTTTGGCAATATCTCTTTTGGAAAGCAGATAGTTTGTTTG	95355							
Qy	153	CATGTATATAAATACAGTCTGATCACACAACAAATTGATGATTGAAAACTACTCAT	212							
Db	95356	AAATCCAAAACAGATTGTTTATCATCATGNTACATTAAITGTTAGGATACATACATACA	95415							
Qy	213	TTAACTTTTATAGAAACCCAAATTAATAGCGGACAACTTTTATAACA	259							
Db	95416	TTAAGTCCTTAGGAATGCAAAAGATTATTGGAAAAAATATATATATA	95462							

RESULT 15
US-09-949-016-16369/c
; Sequence 16369, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949.016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16369
; LENGTH: 136917
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16369

	Query Match	8.2%	Score 41.2;	DB 4;	Length 136917;
	Best Local Similarity	54.7%;	Pred. No. 2.8;		
	Matches	82;	Conservative 0;	Mismatches 68;	Indels 0; Gaps 0;
Qy	219	TTATAGAAACCCAAATTATAGAGCGACAACTTTATAACATGTCAAACTTCGAAGTTAA	278		
Db	93465	TTATTTTATGACAGAAATATAAAACAGTACCANTAGCAGGTATCCAAACATGTAGGAAA	93406		
Qy	279	AAATTTAAGACCCCATAAATTTTACAATTTATAGATTTTAACTCTCCAACTATTTTGTGATGT	338		
Db	93405	AAGAGGNGAAACCATCGCTTTATATAATTACATTTATTTTCGTAACTCAAATTTGCTTTACATT	93346		
Qy	339	TAAAGAAGATACCGAGTCTTTTCTTTCTCA	368		
Db	93345	TGTTAAAAAACTCTAAGTATTTTCTTAAACCA	93316		

Search completed: July 12, 2005, 10:17:50
Job time : 124.576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 543.528 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929C-1_COPY_9829_10329
Perfect score: 501
Sequence: 1 99gtcttagttattgtttt.....gtgattgaattgaacaatt 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	57.8	11.5	5690	14	US-10-239-676-73 Sequence 73, Appl
c 2	57.8	11.5	5690	15	US-10-240-453-79 Sequence 79, Appl
c 3	57	11.4	5153	15	US-10-311-455-928 Sequence 928, App
c 4	57	11.4	10329	15	US-10-311-455-2096 Sequence 2096, App
c 5	53.6	10.7	18512	15	US-10-311-455-950 Sequence 950, App
c 6	53.2	10.6	3673778	16	US-10-312-841-2 Sequence 2, Appl
c 7	51.2	10.2	392	10	US-09-814-353-17543 Sequence 17543, A

8	51.2	10.2	13427	15	US-10-311-455-1900 Sequence 1900, Ap
c 9	51	10.2	3673778	16	US-10-312-841-1 Sequence 1, Appl
c 10	50.6	10.1	6591	15	US-10-311-455-1223 Sequence 1223, Ap
c 11	50	10.0	393	9	US-09-960-352-4582 Sequence 4582, App
c 12	50	10.0	7131	17	US-10-221-613-324 Sequence 324, App
c 13	50	10.0	7455	15	US-10-311-455-1731 Sequence 1731, Ap
c 14	49.6	9.9	5739	15	US-10-311-455-691 Sequence 691, App
c 15	49.4	9.9	17137	15	US-10-311-455-163 Sequence 163, App
c 16	49.2	9.8	10467	15	US-10-240-453-328 Sequence 328, App
c 17	49.2	9.8	10710	15	US-10-311-455-865 Sequence 865, App
c 18	49	9.8	6014	17	US-10-221-613-367 Sequence 367, App
c 19	49	9.8	19380	17	US-10-221-613-389 Sequence 389, App
c 20	48.8	9.7	9415	15	US-10-311-455-267 Sequence 267, App
c 21	48.8	9.7	15743	15	US-10-240-453-270 Sequence 270, App
c 22	48.8	9.7	58965	17	US-10-298-122-2 Sequence 2, Appl
c 23	48.6	9.7	5945	15	US-10-311-455-57 Sequence 57, Appl
c 24	48.6	9.7	6132	15	US-10-311-455-836 Sequence 836, App
c 25	48.6	9.7	6731	17	US-10-257-166-44 Sequence 44, Appl
c 26	48.6	9.7	10891	15	US-10-311-455-437 Sequence 437, App
c 27	48.4	9.7	5576	15	US-10-311-455-1490 Sequence 1490, Ap
c 28	48.4	9.7	5576	15	US-10-240-453-136 Sequence 136, App
c 29	48.4	9.7	6464	15	US-10-311-455-487 Sequence 487, App
c 30	48.4	9.7	9095	19	US-10-433-793-91 Sequence 91, Appl
c 31	48.4	9.7	13420	15	US-10-311-455-890 Sequence 890, App
c 32	48.2	9.6	561	19	US-10-021-323-1540 Sequence 1540, Ap
c 33	48.2	9.6	6963	15	US-10-311-455-496 Sequence 496, App
c 34	47.8	9.5	420	10	US-09-814-353-12772 Sequence 12772, A
c 35	47.8	9.5	421	10	US-09-814-353-5169 Sequence 5169, Ap
c 36	47.4	9.5	5979	14	US-10-239-676-18 Sequence 18, Appl
c 37	47.4	9.5	5979	15	US-10-240-453-26 Sequence 26, Appl
c 38	47.4	9.5	6131	15	US-10-311-455-864 Sequence 864, App
c 39	47.4	9.5	7522	15	US-10-311-455-887 Sequence 887, App
c 40	47.4	9.5	11745	15	US-10-240-453-206 Sequence 206, App
c 41	47.4	9.5	13420	15	US-10-311-455-889 Sequence 889, App
c 42	47.4	9.5	18988	15	US-10-311-455-674 Sequence 674, App
c 43	47.4	9.5	18988	15	US-10-240-485-62 Sequence 62, Appl
c 44	47.4	9.5	18988	18	US-10-221-714A-64 Sequence 64, Appl
c 45	47.2	9.4	6183	15	US-10-311-455-1169 Sequence 1169, Ap

ALIGNMENTS

RESULT 1

US-10-239-676-73/c
; Sequence 73, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 73
; LENGTH: 5690
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:

```
/ NAME/KEY: unsure
/ LOCATION: (381, 706, 711, 1115, 1134, 1206, 1226, 1370, 1382, 1390, 2064)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (2069, 2295, 2337, 2352, 2354, 2361, 2555, 2671, 3057, 3171)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (3186, 3188, 3522, 3568, 3584, 3586, 3591, 3616, 3840, 3846)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (3912, 3999, 4319, 4577, 4600, 4626, 4629, 5299, 5306)
US-10-239-676-73

Query Match      11.5%; Score 57.8; DB 14; Length 5690;
Best Local Similarity 49.2%; Pred. No. 0.031;
Matches 146; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 162 AATACAGTCTGATCACAACAATTTGATGATTTGAAATCTACTGATTTAACTTTA 221
Db 4716 AACGAAACGCGAACCCAACTTAAATAATAATAAATAACAATAAATAACGAAACTA 4657

QY 222 TAGAAACCCAAATTTATAGAGCGCAACTTTATACATGTCGAACTTCGAACTTAAAT 281
Db 4656 ACGAAATAATTTAACGAAACCTTAAANTANTTTCAAACTATCCAACTTCGACTATNAAT 4597

QY 282 TTAAGACCCCATATTTTACAATTTATAGATTTTAACTCTCAACTATTTTGTGATGTTAA 341
Db 4596 AAAAAATCCCTCACTCTAANACTTAAATATACAAATCTTTTATAATTAACCTTTCAAT 4537

QY 342 AGAAGTATCCGAGTCTTTTCTTCAGTTTCCGACCGTCCGACGACTCCCGCCAGCAG 401
Db 4536 TAAAAAATACTATTTCTCTCATCTTCCCTACCCCTACCTATATAAATTTCCGATAAAA 4477

QY 402 TAGAAACGCCAAAGTAAACAAAAGTCGTTAAAAAGTTAAATTAATAAAAAA 458
Db 4476 CAAAAAATACTCAAAACTAACTCAAAATAAAAAACGAAATTAATAATAAAAAA 4420

RESULT 2
US-10-240-453-79/c
; Sequence 79, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 79
; LENGTH: 5690
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (381, 706, 711, 1115, 1134, 1206, 1226, 1370, 1382, 1390, 2064)
```

```
/ NAME/KEY: unsure
/ LOCATION: (2069, 2295, 2337, 2352, 2354, 2361, 2555, 2671, 3057, 3171)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (3186, 3188, 3522, 3568, 3584, 3586, 3591, 3616, 3840, 3846)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (3912, 3999, 4319, 4577, 4600, 4626, 4629, 5299, 5306)
US-10-240-453-79

Query Match      11.5%; Score 57.8; DB 15; Length 5690;
Best Local Similarity 49.2%; Pred. No. 0.031;
Matches 146; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 162 AATACAGTCTGATCACAACAATTTGATGATTTGAAATCTACTGATTTAACTTTA 221
Db 4716 AACGAAACGCGAACCCAACTTAAATAATAATAAATAACAATAAATAACGAAACTA 4657

QY 222 TAGAAACCCAAATTTATAGAGCGCAACTTTTATAACATGTCGAACTTCGAACTTAAAT 281
Db 4656 ACGAAATAATTTAACGAAACCTTAAANTANTTTCAAACTATCCAACTTCGACTATNAAT 4597

QY 282 TTAAGACCCCATATTTTACAATTTATAGATTTTAACTCTCAACTATTTTGTGATGTTAA 341
Db 4596 AAAAAATCCCTCACTCTAANACTTAAATATACAAATCTTTTATAATTAACCTTTCAAT 4537

QY 342 AAGAAGTATCCGAGTCTTTTCTTCAGTTTCCGACCGTCCGACGACTCCCGCCAGCAG 401
Db 4536 TAAAAAATACTATTTCTCTCATCTTCCCTACCCCTACCTATATAAATTTCCGATAAAA 4477

QY 402 TAGAAACGCCAAAGTAAACAAAAGTCGTTAAAAAGTTAAATTAATAAAAAA 458
Db 4476 CAAAAAATACTCAAAACTAACTCAAAATAAAAAACGAAATTAATAATAAAAAA 4420

RESULT 3
US-10-311-455-928/c
; Sequence 928, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 928
; LENGTH: 5153
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-928

Query Match      11.4%; Score 57; DB 15; Length 5153;
Best Local Similarity 48.9%; Pred. No. 0.044;
Matches 153; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 158 ATATAAATACAGTCTGATCACAACAATAATTTGATGATTTGAAATCTACTGATTTAACT 217
Db 4931 ACAAAAATATAAATCAATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4872

QY 218 TTTATAGAAAAACCCAAATTTATAGAGCGCAACTTTTATAACATGTCGAACTTCAAGTTA 277
```

Db 4871 ACTATCTAATCTTTTAAACAACCTTAACAAAACAAAAAATAAAAAAATCCCTATTTA 4812
Qy 278 AAATTTAAGACCCCATTAATTTTACAATTTATAGATTTTAACTCCAACTATTTTGTGATG 337
Db 4811 ATAAATAATACTAAATAAACTAACTAACCATATATAATAAACTAAATAATTCCTTCC 4752
Qy 338 TTAAGAAGATATCCGAGTCTTTTCTTTCGATTTTCCCGACCGTCCCATGACTCCCCAG 397
Db 4751 TTACACCTTATATAAAATTAATTCAAAATAAATTAATAAACTTACATATTAAACCTAAA 4692
Qy 398 CCAGTAGAAAAGCCAAAAGAGTAAACAAAAGCGTTTAAAGAGTTAAATTAATAAAAA 457
Db 4691 CCATAAAACCTTAAAAAAAACCTTAAACATATACCATTCAAAACATAAACATAACAAA 4632
Qy 458 AATAGATAGTTGA 470
Db 4631 ACTTCATATCTAA 4619

RESULT 4

US-10-311-455-2096/c
; Sequence 2096, Application US/10311455
; Publication No. US20030143606A1

; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2096
; LENGTH: 10329
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-2096

Query Match 11.4%; Score 57; DB 15; Length 10329;
Best Local Similarity 48.9%; Pred. No. 0.056;
Matches 153; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
Qy 158 ATATAATACAGTCTGATCACAACAACAAATTTGATGATTTGAAATTAATACTACTGATTTAAC 217
Db 8678 ACAAATAACAACCAATAAACAACAACAAACCCCTCAAAATAATACCAATATCTACA 8619
Qy 218 TTTATAGAAAACCAAAATTTAGAGCGACAACTTTTATAAATCATGTCGAAAGTTA 277
Db 8618 ACTATCTAATCTTTAAACCTTAACAACAACAACAATAAAAAAATTCCTATTTA 8559
Qy 278 AAATTTAAGACCCCATTAATTTTACAATTTATAGATTTTAACTCCAACTATTTGTGATG 337
Db 8558 ATAAATAATACTAAAAAATACTAACCATATATAAAAAAATAAACTAAATCCCTTCC 8499
Qy 338 TTAAGAAGATATCCGAGTCTTTTCTTTCGATTTTCCCGACCGTCCCATGACTCCCCAG 397
Db 8498 TTATACCTTATACAAAATTAATTCAAAATAAATTAATAAACTTCAATATTAAACCTTAAA 8439
Qy 398 CCAGTAGAAAAGCCAAAAGTAAACAAAAGCGTTTAAAGAGTTAAATTAATAAAAA 457
Db 8438 CCATAAAACCTTAAAAAAAACCTTAAACATATACCATTCAAAACATAAACATAACAAA 8379
Qy 458 AATAGATAGTTGA 470

Db 8378 ACTTCATATCTAA 8366

RESULT 5

US-10-311-455-950/c
; Sequence 950, Application US/10311455
; Publication No. US20030143606A1

; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 950
; LENGTH: 18512
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-950

Query Match 10.7%; Score 53.6; DB 15; Length 18512;
Best Local Similarity 47.6%; Pred. No. 0.36;
Matches 158; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
Qy 129 TTTTGAAAACCAACGTTCTACTTTTCATGTATATAAATACAGTCTGATCACAACAACAAA 188
Db 13264 TTTAAAAATAATTTCTTTTATATATATATATATACAAAAACATTTAAACAAAAATTT 13205
Qy 189 TTGATGATTTGAAATTAATACTACTGATTTTAACTTTATAGAAAACCCAAATTTATAGCGCAA 248
Db 13204 TTACTATACAAAACCTATTAATTTACTATATAATTAATAATCAAAAATTTTACCAAACT 13145
Qy 249 CTTTATAAACATGTCAAAACCTTCGAAGTTTAAATTTTAAAGACCCCATATTTTACAATTTATA 308
Db 13144 CCTCCTATATATAAATAACAATACTACGAAAACAATAAACTAATCTCTTCTTCAACAA 13085
Qy 309 GATTTTAAATACTCCAATATTTTGTGATGTTTAAAGAGTATCCGAGTCTTTTCTTTTCCA 368
Db 13084 CCTCTTAATACGCTCTATATATAAAAAAATAAATAAACAATTTTAAACATCTA 13025
Qy 369 GTTTTCCCGCCGTCCTGACTCCCGGACCTAGAAAAGCCAAAAGTAAACAAA 428
Db 13024 AAAACAATAAATACAAATTTACAAAACAATATATATTTTAAATATCAAAAAATAATACAAAT 12965
Qy 429 AGTCGTTTAAAAAGTTTAAATTTAAAAAATAAAT 460
Db 12964 ATATAATAAAAAACTTCAATAAAAAACAAT 12933

RESULT 6

US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1

; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des Menschen
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312.841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2

```

; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
; US-10-312-841-2

Query Match      10.6%;      Score 53.2;  DB 16;      Length 3673778;
Best Local Similarity 48.7%;      Pred. No. 2.7;
Matches 145; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy      158  ATATAAATACAGCTCGATCACAACAACAATTTGATGATTTGAAAATACTACTGATTTAAAC 217
Db      576885  ATATAAAAATTAAAAAANAACACACAAAAAATAATTTCAAAAAAACAATTAATCACTTAAA 576826

Qy      218  TTTATAGAAAAACCCAAATTTATAGAGCGCAACTTTTATAAACATGTCAAACTTTCGAAGTTA 277
Db      576825  TACATTAATACCTAAAAATAAACATTTATTAACTTTTAATATCTCAATAACCAAAATAA 576766

Qy      278  AAATTTAAGACCCCAATAATTTTACAAATTATAGATTTTAAATAGTCCAACTATTTTGTGATG 337
Db      576765  CAATAAAACAATTTAAAAATAAACAATAAAAAAATAACAAATTATAAAAAAACTTTAAATATT 576706

Qy      338  TTTAAAGAAGTATCCGAGTCTTTCTTTCCAGTTTCCCCACCGTCCCACTGACTCCCCAC 397
Db      576705  TCAAAAACAATAATTTTATTTATTTATTTATTTTAAAAATAATTTAAAAAATTAATTA 576646

Qy      398  CCAGTAGAAAAAGGCCAAAAAAGTAAACAAAAAGTCGTTAAAAAAGCTTAAATTTAAAAA 455
Db      576645  ATCAACACAAACAATCAAAAAATTATCAAAATATTAAATAAAAAACAATCTTAAATAAAAA 576588

```

```

RESULT 7
US-09-814-353-17543/c
; Sequence 17543, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17543
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218,
; LOCATION: 219, 220, 221, 222, 243, 244, 245, 250, 253, 288, 289, 290,
; LOCATION: 292, 293, 294, 295, 316, 317, 335, 337, 341, 363
; OTHER INFORMATION: n = A.T.C or G

```

```

US-09-814-353--17543

Query Match          10.2%; Score 51.2; DB 10; Length 392;
Best Local Similarity 44.7%; Pred. No. 0.32;
Matches 119; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 234 ATTATAGAGCGACAACTTTTATAACATGTGCAAACTTCGAAGTTAAAAATTTAAGACCCCAT 293
Db 364 ANAATTAATAAAATATATTTTTTCNCAANTWAAAAAATAAAATAAAANNAATTTATTTTAT 305

Qy 294 AATTTTACAATATAGATTTTAAATCTCCAACACTATTTTGTGATGTGTAAAAAGAGATATCCG 353
Db 304 TTTTTTTTTNNNNNTNNAAAAAATTTTTTTTTTTTAAAAAANAANAATATN 245

Qy 354 AGTCCTTTCTTCCAGTTTCCCCACCGTCCCATGACTCCCCGACGAGTAGAAAAAGCA 413
Db 244 NNTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNNNNCATCAAAAAAATAAAAAAAA 185

Qy 414 AAAAAAGTAAACAAAAAGTCGTTAAAAAGTTAAATTTAAAAAATAAGATAGTAGTGCAGT 473
Db 184 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 125

Qy 474 TTACTAAAGTGATTTGAAATTGAACAA 499
Db 124 AAAAAAATAAAAAAATAAAAAATTTTAAAAAA 99

```

RESISTANCE

```

RES001.8
US-10-311-455-1900
; Sequence 1900, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1900
; LENGTH: 13427
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-311-455-1900

```

	Query Match	10.2%; Score 51.2; DB 15; Length 13427;
	Best Local Similarity	50.0%; Pred.No. 1.1; Indels 5; Gaps 2;
	Matches 183; Conservative 0; Mismatches	178;
QY	4	TCTTAGTATTGATTTTTAGAGAGTTAATAATAGTATATATATATATATATGTATATACTAGCATTGGT 63
Dd	567	TGTTATTAGTATTTATTCGGAGATTAAATATGTTATTATTATTTTGTTTTTAAAAAGTT 626
QY	64	ACCTAATCTCTGTCGTTTAATCTCAATTT --- AGTGAGTCAGGAACCAGACTCGTTGGGTAA 119
Dd	627	TTTATAGGCCGGGNTTAAAAGTTAGTTTAAITTAGTGAATTTTITAGATTGGAG 686
QY	120	GGCCTCTCCCTTTTGAACCGAACGTTCTCATCTGTATATAATACAGCTCTGATCAC 179
Dd	687	AATATTGGTTTAAAAATTTTGGAAATTTTTTTTTTAGGTAAANAATGGAATGTTGAAGTCG 746
QY	180	C AACACAAATTTGATGATTGAAAATFACTGATTTAACTTTTATAGA AAA ACCCAAATTTATA 239

Db 747 GAAGA-ATAATTGATGTTTGATTAATTTTGAAGTATATATATATATATATATATATATA 805
Qy 240 GAGCGACAACCTTTATAAACAATGTCAAACTTCGAAGTTTAAATTTAAGACCCCAATAATTTT 299
Db 806 TATATATATATATATATATATATACGAATTCATAGATGGAATAATATGTAATAAATG 865
Qy 300 ACAATTATAGATTTTATATACCCAACTATTTTGTGATGTTTAAAGAAGTATCCGAGTCCT 359
Db 866 TTAATCTAGAAATTTAGGTGGTGAATTTTGGGGTGTATTATATATAGAAATTTTAAATTTG 925
Qy 360 TTCTTTT 365
Db 926 TTTT 931

RESULT 9
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; PRIORITY FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 10.2%; Score 51; DB 16; Length 3673778;
Best Local Similarity 47.9%; Pred. No. 7.8;
Matches 147; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
Qy 187 AATTGATGATGAATACTACTGATTTAACTTTATAGAAACCCCAATTTATAGAGGCAC 246
Db 410521 ATTAAATTAATACTAATTTTACTACCTAACTCTCAATAATAAATAAATCTCTATATCTTTATTC 410462
Qy 247 AACTTTATAAACAATGTCAAACTTCGAAGTTTAAATTTTAAAGACCCCAATAATTTTACAAATTA 306
Db 410461 ACTATATAATAAATACCTAAATTTATACCTATACATATAATAATAATATCCATTAATAA 410402
Qy 307 TAGATTTTAACTATCCAACTATTTTGTGATGTTTAAAGAAGTATCCGAGTCCTTTCTTTTC 366
Db 410401 AACAAATAAATTTATCTTCTTAATTTCTTTCTCTAACTATATATATCTTTTAAATTAATCTTC 410342
Qy 367 CAGTTTCCCAACCGTCCCATGATCTCCCAAGCAGTAGAAAGACCAAAAGTAAACAA 426
Db 410341 TTATATAACCATTAACAAATTTTATACCAACAACACAAATAATCAATAAATCAAAACCT 410282
Qy 427 AAGTCGTTTAAAGAAGTTTAAATTTAAATAAATAAATAAATAGATGTCGCTTTTACTAAGTAT 486
Db 410281 AATTACAACAAATAAATAAATAAATTTAATCGTAAACCCCAATAAATAAATAAATAAATAA 410222
Qy 487 TTGAAT 493
Db 410221 TTCAAT 410215

RESULT 10
US-10-311-455-1223/c
; Sequence 1223, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter.
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIORITY FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIORITY FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIORITY FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIORITY FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1223
; LENGTH: 6591
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1223

Query Match 10.1%; Score 50.6; DB 15; Length 6591;
Best Local Similarity 46.1%; Pred. No. 1.1;
Matches 206; Conservative 0; Mismatches 239; Indels 2; Gaps 1;
Qy 54 TAAAGTAGGTACCTAATCTCTGCTTAATCTAATTTTAGTGAGTCAGGAACCGACTCGTT 113
Db 2843 TAAACATTTCTCTTTATCGTCTTAATTTTATAAAACTAACTTTCTTTTAAACTATAT 2784
Qy 114 GGCTAAGGTCTCTCTTTTGAAGCGCAAGCTTCTACTTTTCATGTATATAATACAGTCGT 173
Db 2783 CTCTTCGATCCCTCATCTACTCTTCTCTATCCATCTCTCCATCTATTTCTAATTTCT 2724
Qy 174 ATCACACAACAAATTTGATGTTGAAATATCTACTGATTTAACTTTTATAGAAACCCAA 233
Db 2723 TCTACAAATCCCATTTATATATAATAATAATAATAATAATAATAATAATAATAATA 2664
Qy 234 ATTATAGAGCGACAACCTTTTATAAACAATGTCAAACTTCGAAGTTTAAATTTTAAAGACCCAT 293
Db 2663 ACTCTTCTTTCTCTCTCTATCACTTCTCTCCCACTTTTCAAAACAACCTTTTATAAATPAC 2604
Qy 294 AATTTTACAATTAAGATTTTAACTACTCCAACTATTTTGTGATGTTTAAAGAAGTATCCG 353
Db 2603 AA--CTCCAATCATATATTTTATAATAAATACTTAAACTTAACGCATATAAACAATAA 2546
Qy 354 AGTCCTTTCTTCCAGTTTCCCAACCGTCCCACTGATCTCCCAAGCAGTAGAATAAGCCCA 413
Db 2545 AACTTTTACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2486
Qy 414 AAAAAAGTAAACAAAAAGTCGTTTAAAAAAGTTTAAATTTAAAAAATAAATAAATAAATAA 473
Db 2485 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2426
Qy 474 TTACTAAGTGAATTCGAATTTGAACAAT 500
Db 2425 TTACTAATAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2399

RESULT 11
US-09-960-352-4582/c
; Sequence 4582, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIORITY FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112

```
; SEQ ID NO 4582  
; LENGTH: 393  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7  
US-09-960-352-4582
```

```
Query Match      10.0%; Score 50; DB 9; Length 393;  
Best Local Similarity 49.2%; Pred.No. 0.57;  
Matches 131; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
```

Qy	200	AATACTACGTATTTACTTTATAGAAAAAACCCAAAATTATAGGCGCACACTTATAAACCA	259
Db	393	AAATTAAAIAAAAAAAAAATTAATTTAAAAAATTAATATAAAAAAAAAAATTTATAAAAA	334
Qy	260	TGTCAAACTTCGAAGTTAAAATTTAAGNCCCATAAATTTACAATTTATAGATTTTTATATC	319
Db	333	TATTTAAATTTTATAATAATAAAATAATATAATCAAATAATAATAATAATTTATAATA	274
Qy	320	TCCAACATATTTTGFGATGTTTAAAAAGAAGTATCCGAGTCTTTCTTCCAGTTCCCACC	379
Db	273	ATTTTTTAAAAAATAATATATTTTAAAAAIIIIIIIIIIITATAAATAATTTTTTAAA	214
Qy	380	GTCGCCATGACTCCCCGCAGTAGAAAAAGCCAAAAAGTAAACAAAAAGTCGTTAAAA	439
Db	213	ATAAAAAAATTTATAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	154
Qy	440	AAGTTAAATTTAAAAAAATAAGATA	465
Db	153	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	128

```

RESULT 12
US-10-221-613-324/c
; Sequence 324, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221.613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 324
; LENGTH: 7131
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-324

```

Query Match 10.0%; Score 50; DB 17; Length 7131;
Best Local Similarity 46.5%; Pred. No. 1.5;
Matches 161; Conservative 0; Mismatches 185; Indels 0; Gaps 0;
QY 150 TTTTCATGCTATATAAATACAGTCTCATACACACACAAATTTGATGATTTAAAAATACCTACT 209
Db 5741 TTTAAAAATCATTTAACTTTATATAAACTTTAAAATTTTATTTTATTTTCTATTTAAAAAAATTTTA 5682

Qy 210 GATTTAACCTTTATAGAAAACCCAAATATATAGCGGCAACACTTTTATAAACATGTCAAACTT 269

Db 5681 ACTACAAAAAATATAAAACCAATCTAAATAAATTTAAATCATAAAAACACTTTAAACACTT 5622

Qy 270 CGAAGTTAAAAATTTAAGACCCCAATAATTTTACAAATATATAGATTTTAATCTCCAAACTATT 329

Db 5621 CATATAATAATATAAAATATATATTAATCTCAAAATTTAAATTAATTAATTCATTAATA 5562

Qy 330 TTGTGATGTTTAAAGAAGTATCGAGTCTTTTTCTTTCCAGTGTTCCTCCACCGTCCCATGAC 389

Db 5561 ACCTCAAAAACCTCAAAATCTCTCTATCTCTCTTTTACCATTCTTTTAAATCGATTTTA 5502

Qy 390 TCCCCGAGCAGTAGAAAAGCCAAAAGAGTAAACAAAAGTCGTTAAAAAAGTTAAATT 445

Db 5501 TTAACACGTAATTTATATATAAATTCAAATTTTAAATATATACTAAAAAAATTTTAAACAA 5442

Qy 450 AAAAAAATAATAGATAGTTGACGTTTACTAAAGTGAATTCGAATTGA 495

Db 5441 TATATAATAATATCTATTACCACAATAAAAAAATTTAACATTTA 5396

RESULT 13

US-10-311-455-1731/c

; Sequence 1731, Application US/10311455

; Publication No. US20030143606A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System

; TITLE OF INVENTION: cytosine methylation

; FILE REFERENCE: 5013.1014

; CURRENT APPLICATION NUMBER: US/10/311,455

; PRIOR FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537

; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 2424

; SEQ ID NO 1731

; LENGTH: 7455

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-1731

	Query Match	10.0%;	Score 50;	DB 15;	Length 7455;
	Best Local Similarity	48.5%;	Pred. No. 1.6;		
	Matches 167;	Conservative 0;	Mismatches 175;	Indels 2;	Gaps 1;
Qy	140	AAGCTTCTACTTTTCATGTATATAAATACAGTCTCGATCACACACACAAATTGATGATTGA	199		
Db	1248	AAACATTTTATTAAAAAATAAACTATAAACTTAAATACAAAAATATTATTATTA	1189		
Qy	200	AAATACTACTGATTTTAACTTTTATAGAAAAACCCAAATTTATAGAGCGCAACTTTTATAAACA	259		
Db	1188	CTTAACAAATTTATATACCACTTTATAAACATATAAATAAATCTTTTCTTTAT--ACA	1131		
Qy	260	TGTCAAACTTTCGAAGTTTAAAAATTTAAGACCCCATAAATTTTACAATTTAGATTTTAAATAC	319		
Db	1130	TTTACGTTTTCTTAAATAAAATAACAATATAAAAAATTTAACAATTAATAAAATTTAAAT	1071		
Qy	320	TCCAACATTTTGTGATGTTTAAAGAAGTATCCGAGTCTTTTCTTTCCAGTTTCCCCACC	379		
Db	1070	TTTAAACAAATATAAATTAATCAAAATTCAAACTTCACTAAAAATTTATTTAAATTTATTTAAAT	1011		
Qy	380	GTCCCATGACTCCCCGACGATAGAAAAAGCCAAAAAGCTTAAACAAAAAGTCGTTAAAAA	439		
Db	1010	ATATACTAAATCATAAATATTTAAATTTATAAATTTAAAAAATATATAAAAAAATAAAATTAAC	951		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:03:44 ; Search time 2468.64 Seconds
(without alignments)
7724.974 Million cell updates/sec

Title: US-10-018-929c-1_COPY_9829_10329

Perfect score: 501

Sequence: 1 gggcttagtattgatttt.....gtgattgaattgaacaatt 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hcc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gss1.*

9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403.8	80.6	566	8 B67281	B67281 T22N9TR TAM
2	332.6	66.4	472	8 B62563	B62563 T22D24TR TA
3	259.6	51.8	639	8 B20434	B20434 T22N9-Sp6 T
4	219.4	43.8	661	8 B20425	B20425 T22D24-Sp6
5	196.4	39.2	205	1 AA597930	AA597930 29179 lam
6	120.6	24.1	398	1 AU230038	AU230038 AU230038
7	62.4	12.5	1101	9 CNS000EVL	AL069706 Drosophil
8	58	11.6	996	9 CNS000FUH	AL071063 Drosophil
9	57.6	11.5	1190	9 CNS00039G	AL063921 Drosophil
10	57.2	11.4	928	9 CNS020N7	AL206908 Tetradon
11	56.6	11.3	741	8 BH586848	AL071865 Drosophil
12	56.6	11.3	1001	9 CNS01400	BH586848 BOGJ24ITF
13	56.4	11.3	843	9 CNS000CS1	AL103554 Drosophil
14	55.4	11.1	1044	9 CNS0003IX	AL059666 Drosophil
15	55.4	11.0	1043	9 CNS00145P	AL064565 Drosophil
16	55.2	10.7	987	9 CNS014PQ	AL103735 Drosophil
17	53.8	10.7	1546	9 CL502023	AL104456 Drosophil
18	53.8	10.7	1546	9 CL502023	CL502023 SAIL_706
19	53.6	10.7	677	8 BH959836	BH959836 odi08h08
20	53.6	10.7	702	8 BH959856	BH959856 odi08h08
21	53.4	10.7	668	8 BH551540	BH551540 BOGTRB2TF
22	53.2	10.5	878	9 CNS0187R	AL108993 Drosophil
23	52.8	10.5	1015	9 CNS000IUO	AL075463 Drosophil
24	52.8	10.5	1179	9 CNS06Q7Y	AL410324 T7 end of

25	52.8	10.5	1200	9 CNS016CO	AL106578 Drosophil
26	52.6	10.5	994	9 CNS04NOJ	AL298972 Tetradon
27	52.4	10.5	759	9 CNS06QXV	AL411257 T7 end of
28	52.4	10.5	893	9 CNS013XE	AL103436 Drosophil
29	52.4	10.5	1020	4 BM415405	BM415405 OP20479 M
30	52.4	10.5	1391	9 CG754863	CG754863 P050-2-G0
31	52	10.4	499	8 AZ030520	AZ030520 RPCI-23-2
32	52	10.4	604	8 AQ925473	AQ925473 RPCI-23-2
33	52	10.4	821	9 CNS03F7F	AL241332 Tetradon
34	52	10.4	1101	9 CNS00EO7	AL069440 Drosophil
35	51.8	10.3	945	9 CNS04D0K	AL285149 Tetradon
36	51.8	10.3	1101	9 CNS0039G	AL063921 Drosophil
37	51.8	10.3	1101	9 CNS007D4	AL067015 Drosophil
38	51.6	10.3	855	9 CNS04P5D	AL300874 Tetradon
39	51.6	10.3	886	8 BH177277	BH177277 008_L_22-
40	51.6	10.3	886	9 CNS07JUX	AL614235 T3 end of
41	51.4	10.3	361	6 CB317278	CB317278 AGENCOURT
42	51.2	10.2	725	8 AQ273192	AQ273192 nbxb0029K
43	51.2	10.2	1255	9 AG435644	AG435644 Mus muscu
44	51.2	10.2	2158	3 BC039314	BC039314 Homo sapi
45	51	10.2	706	8 CC075164	CC075164 CSU-K33r.

ALIGNMENTS

RESULT 1

B67281

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSEs: T22N9TF

Contact: Steve Rounsley

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: rounsley@tigr.org

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 566.

Location/Qualifiers

1..566

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Columbia"

/db_xref="taxon:3702"

/clone="T22N9"

/sex="hermaphrodite"

/clone_lib="TAMU"

/notes="Vector: BelOBACII; Site 1: HindIII; Site 2: HindIII; Produced by Rod Wing"

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Score

DB

Length

Indels

Gaps


```

/notes="Vector: BelobACII; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"

ORIGIN
Query Match      51.8%; Score 259.6; DB 8; Length 639;
Best Local Similarity 78.1%; Pred. No. 5.2e-43;
Matches 336; Conservative 0; Mismatches 91; Indels 3; Gaps 2;

QY 1 GGGCTCTAGTATGATTTAGAGAGTTAAATAATAGTATATATATATATATATAGTAG 60
    |||||
Db 198 GGGCTCTAGTATGATTTAGAGAGTTAAATAATAGTATATATATATATATATAGTAG 257

QY 61 GTTACCTAACTCTGTCGTTAACTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAG 120
    |||||
Db 258 GTCACTTACTCTCTGTCGTTAACTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAG 317

QY 121 GTCTCTCTTTTGAACGCAACGTTCTACATTTTCAATGATATATAATACAGTCTGATACAC 180
    |||||
Db 318 GTCTCTCTTTTGAACGCAACGTTCTACATTTTCAATGATATATAATACAGTCTGATACAC 377

QY 181 AACACAAATTTGATGATTTGAAATACTACTGATTTAACTTTTATAGAAACCCAAATATAG 240
    |||||
Db 378 AACACAAATTTGATGATTTGAAACATCTACTGATTTAACTTTTATAGAACCCAAATATAG 437

QY 241 AGGACAACTTTTATAAATGTCGAAGTTAAATTTTAAAGACCCCAATATTTTA 300
    |||||
Db 438 AGGACAACTTTTATAAATGTCGAAGTTAAATTTTAAAGACCCCAATATTTTA 495

QY 301 CAATTAT-AGATTTTAACTCCAACTATTTTGTGATTTTAAAGAGTATCCGAGTCTT 359
    |||||
Db 496 CATTAAGTATTTAACTCCAACTATTTTGTGATTTTAAAGAGTATCCGAGTCTT 555

QY 360 TTCTTTCCAGTTTCCCAACCGTCCCATGACTCCCAAGCAGTAGAAGCAACCAAGTCTTAAAG 419
    |||||
Db 556 TTCCCAACCGTCCCAACCGTCCCAAGCAGTAGAAGCAACCAAGTCTTAAAG 615

QY 420 TAAACAAAAA 429
    |||||
Db 616 TAAATTAATA 625

RESULT 4
B20425
LOCUS
DEFINITION
T22D24-Sp6 TAMU Arabidopsis thaliana genomic clone T22D24, genomic
survey sequence.
ACCESSION
B20425
VERSION
B20425.1 GI:2395479
SOURCE
GSS.
ORGANISM
Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
AUTHORS
BAC End Sequences at ATGC
TITLE
Unpublished (1997)
JOURNAL
Other_GSSs: T22D24-T7
COMMENT
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@genome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 268
High quality sequence stop: 300.
Location/Qualifiers

ORIGIN
Query Match      43.8%; Score 219.4; DB 8; Length 661;
Best Local Similarity 80.8%; Pred. No. 8.4e-35;
Matches 256; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 GGGCTCTAGTATGATTTAGAGAGTTAAATAATAGTATATATATATATATATAGTAG 60
    |||||
Db 198 GGGCTCTAGTATGATTTAGAGAGTTAAATAATAGTATATCTCTCTATCTCTCTAGTCG 257

QY 61 GTTACCTAACTCTGTCGTTAACTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAG 120
    |||||
Db 258 GTCCCTTACTCTGTCGTTAACTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAG 317

QY 121 GTCTCTCTTTTGAACGCAACGTTCTACTTTTCTATGATATATAATACAGTCTGATACAC 180
    |||||
Db 318 GTCTCTCTTTTGAACGCAACGTTCTACTTTTCTATGATATATAATACAGTCTGATACAC 377

QY 181 AACACAAATTTGATGATTTGAAATACTACTGATTTAACTTTTATAGAAACCCAAATATAG 240
    |||||
Db 378 AACACAAATTTGATGATTTGAAACATCTACTGATTTAACTTTTATAGAACCCAAATATAG 437

QY 241 AGGACAACTTTTATAAATGTCGAAGTTAAATTTTAAAGACCCCAATATTTTA 300
    |||||
Db 438 AGGACAACTTTTATAAATGTCGAAGTTAAATTTTAAAGACCCCAATATTTTA 497

QY 301 CAATTATAGATTTTAAAT 317
    |||||
Db 498 CACTTATGATCTTACT 514

RESULT 5
AA597930
LOCUS
DEFINITION
29179 Lambda-PRL2 Arabidopsis thaliana cDNA clone 247N2T7, mRNA
sequence.
ACCESSION
AA597930
VERSION
AA597930.1 GI:2413353
SOURCE
EST.
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 205)
Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H.,
McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M.,
Retzel,E. and Somerville,C.
AUTHORS
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
PUBMED
7846151
COMMENT
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcm@ibm.cl.msu.edu
Seq primer: T7 dye primer.
Location/Qualifiers

FEATURES
```

FEATURES

source	details.	FEATURES	Location/Qualifiers
1. .205		source	1. .398
/organism="Arabidopsis thaliana"			/organism="Arabidopsis thaliana"
/mol_type="mRNA"			/mol_type="mRNA"
/ecotype="Columbia"			/db_xref="taxon:3702"
/db_xref="taxon:3702"			/clone_lib="RAFL19-14-009"
/clone="247N2T7"			/tissue_type="mixture of silique and flower"
/clone_lib="Lambda-PRL2"			/lab_host="DH10B"
/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."			/clone_lib="RAFL19"
			/note="Site_1: BamHI; Site_2: SalI; Subtraction Library"
ORIGIN			
Query Match	39.2%; Score 196.4; DB 1; Length 205;		
Best Local Similarity	96.1%; Pred. No. 4.4e-30;		
Matches	197; Conservative 0; Mismatches 8; Indels 0; Gaps 0;		
QY	27 GTTAATAAGTATATATATATATATAGTATAGTAGTACCTAATCTCTGCTTAATCTA 86		
Db	1 GTTAATAAGTATATATATATATATAGTATAGTAGTACCTAATCTCTGCTTAATCTA 60		
QY	87 ATTAGTAGTCAGGAACCGACTCGTTGGCTAAGGTCCTCTCTTTTGAACGCAACGTTTC 146		
Db	61 ATTAGTAGTCAGGAACCGACTCGTTGGCTAAGGTCCTCTCTTTTGAACGCAACGTTTC 120		
QY	147 TACTTTTCATGTATATAATACAGTCCTGATCACACACAAATTTGATGATTTGAAATATCT 206		
Db	121 TACTTTTCATGTATATAATACAGTCCTGATCACACACAAATTTGATGATTTGAAATATCT 180		
QY	207 ACTGATTAACTTTATAGAAAACCC 231		
Db	181 ACTGATTAACTTTATAGAAAACCC 205		
RESULT 6			
AU230038/c			
LOCUS	398 bp mRNA linear EST 25-APR-2002		
DEFINITION	AU230038 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-14-009 3', mRNA sequence.		
ACCESSION	AU230038.1 GI:19798745		
VERSION	EST.		
KEYWORDS	Arabidopsis thaliana (thale cress)		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 398)		
AUTHORS	Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shingawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.		
TITLE	Large scale analysis of Arabidopsis full-length cDNA		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Motoaki Seki Plant Functional Genomics Research Group RIKEN Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel: 81-298-36-4359 Fax: 81-298-36-9080 Email: mseki@rtc.riken.go.jp An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further		
FEATURES			
source			
1. .1101			
/organism="Drosophila melanogaster"			
/mol_type="genomic DNA"			
/db_xref="taxon:7227"			
/clone="BACR29B23"			
/clone_lib="RPCI-98"			
/note="end : T7"			
ORIGIN			
Query Match	24.1%; Score 120.6; DB 1; Length 398;		
Best Local Similarity	93.0%; Pred. No. 1.3e-14;		
Matches	159; Conservative 0; Mismatches 9; Indels 3; Gaps 3;		
QY	1 GGGTCTTAGTTA-TTGATTTTGTAGAGCTTAATATAGTATATATATATATATATAGTA 59		
Db	178 GGGTCTTAGTTA-TTGATTTTGTAGAGCTTAATATAGTATATATATATATATATAGTAG 119		
QY	60 GGTACCTAATCTCTGTCG-TTAAATCTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTTA 118		
Db	118 GTTACCTAATCTCTGTCG-TTAAATCTAATTTAGTGAGTCAGGAACCGACTCGTT-GCTA 60		
QY	119 AGGTCTCTCTTTTGAACGCAACGTTCTTCTCATGTATATAAATACAG 169		
Db	59 AGGTCTCTCTTTTGAACGCAACGTTCTTCTCATGTATATAAATACAG 9		
RESULT 7			
CNS000EVL	1101 bp DNA linear GSS 04-JUN-1999		
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC:		
DEFINITION	BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL069706		
VERSION	AL069706.1 GI:4949849		
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster (fruit fly)		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)		
COMMENT	- Web : www.genoscope.cns.fr Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw ep, the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .		

AL103554
 AL103554.1 GI:5615165
 GSS.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1001)
 Genoscope.
 Direct Submission
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk -- This Drosophila melanogaster BAC
 library (dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.

FEATURES
 source Location/Qualifiers
 1..1001
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN11B06"
 /clone_lib="drosBAC"
 /plasmid="pBelobAC11"
 /note="end : sp6"

ORIGIN
 Query Match 11.3%; Score 56.6; DB 9; Length 1001;
 Best Local Similarity 21.4%; Pred. No.0.15; Indels 0; Gaps 0;
 Matches 98; Conservative 146; Mismatches 215;

Qy 7 TAGTATTGATTTTGTAGAGAGTAAATAATGATATATATATATATATATAGTAGGTACC 66
 Db 991 YHMTTAAHAHTMTWTAHTTAMTAHTATWTMTCTWMTHTTTATTTTWTMTWHT 932
 Qy 67 TAATCTCTGCGTTAACTTAATTTAGTAGAGTCAGAAACCGACTCGTTGGCTAAGGTCTCT 126
 Db 931 MTHTCMHDTVTWTCWMTCNMTWNTWNTTATATAATHTTTTWTWMTWTTTAHDTMT 872
 Qy 127 CCTTTGAAACGCCAAGCTTCTACTTTTCATGTATATAAATACAGTCTGATCACAAACACA 186
 Db 871 VTTWMTWMTATAAHTTMMWAAHWHATTWTTWCHWTWTTAHTTWTWTTATMTAVTADW 812
 Qy 187 AATTGATGATTGAAATACTACTCATTTTAACCTTTATAGAAACCCAAATATAGAGCGAC 246
 Db 811 DWTNKKATAHTMTAAWTTCTYTYTYTTHCHCTCCTCTCTMYNMMHWHATCCCMKCYAY 752
 Qy 247 AACTTTTAAACATGTCACAACTTCGAGAGTTAAAAATTTAAGACCCCAATAATTTACAATTA 306
 Db 751 TMMMHMMHYMHMTMYMYCCTMTYMNNTWMAAHTWTTWTAACAAHTAAHAHTMTAHTM 692
 Qy 307 TAGATTTTATACTCCCACTATTTTGTGATGTTTAAAGAGAGTATCCGAGTCTTTTCTTC 366
 Db 691 TARTWMTWMTWNCNMCMTWMBYTTWTTTATWTTCTTHAMAAWYMTWTTTTTWTYYAH 632
 Qy 367 CAGTTTCCCAACCGTCCCATGACTCCCGCCAGTAGTAAAGAAAGCCAAAAAAGTAAACAA 426
 Db 631 MWMACYCHMAWNAVATDNGYNTCNTVKVXABITTCYTTASTRANAANAAWNAANATAW 572
 Qy 427 AAAGTCGTTAAAAAGTTAAATTTAAAAAATAAATAGATA 465
 Db 571 TTTAANNNAANAANAANAANAANAANANNANNAANA 533

RESULT 14
 CNS00CS1/c 843 bp DNA linear GSS 04-JUN-1999
 LOCUS

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2529.46 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929C-2_COPY_1_500
Perfect score: 500
Sequence: 1 cacaacagatgagttttctt.....agactggctccctcaactgc 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	6554	8 AF213627	AF213627 Arabidops
2	500	100.0	6571	6 AX063583	AX063583 Sequence
3	304	60.8	10329	6 AX063582	AX063582 Sequence
4	304	60.8	10329	8 AF213628	AF213628 Arabidops
5	304	60.8	120965	8 AC026875	AC026875 Genomic s
6	39.4	7.9	221544	2 BX927290	BX927290 Danio rer
7	38.2	7.6	221873	2 AC096694	AC096694 Rattus no
8	38.2	7.6	225918	2 AC094977	AC094977 Rattus no
9	38.2	7.6	234966	2 AC131437	AC131437 Rattus no
10	37.6	7.5	88383	8 AC006233	AC006233 Arabidops
11	37.6	7.5	90341	8 AC006232	AC006232 Arabidops
12	37.4	7.5	112180	9 HSJ322G13	HSJ322G13 Human DNA
13	37.2	7.4	39329	9 AC008995	AC008995 Homo sapi
14	37.2	7.4	181532	2 AC020700	AC020700 Homo sapi
15	37.2	7.4	184213	9 AC073539	AC073539 Homo sapi
16	37	7.4	1512	3 AF153275	AF153275 Caenorhab
17	37	7.4	1611	3 AB027412	AB027412 Caenorhab
18	37	7.4	130135	2 AC146752	AC146752 Medicago
19	36.8	7.4	131347	2 AC002421	AC002421 Homo sapi

20	36.8	7.4	170623	9	AL391375	AL391375 Human DNA
21	36.4	7.3	140349	9	AC027320	AC027320 Homo sapi
22	36.4	7.3	175485	2	AC073802	AC073802 Mus muscu
23	36.4	7.3	209461	2	AF528163	AF528163 Mus muscu
24	36.4	7.3	220832	2	AC126145	AC126145 Rattus no
25	36.2	7.2	3840	8	SCYGL009C	SCYGL009C
26	36.2	7.2	4607	8	SCYGL008C	SCYGL008C
27	36.2	7.2	140338	5	BX322629	BX322629 Zebrafish
28	36.2	7.2	161672	2	BX276184	BX276184 Danio rer
29	36.2	7.2	215374	2	AC150765	AC150765 Bos tauru
30	36	7.2	144017	2	AC141685	AC141685 Apis mell
31	35.8	7.2	157250	2	AC027264	AC027264 Homo sapi
32	35.8	7.2	173133	10	AL844604	AL844604 Mouse DNA
33	35.8	7.2	190748	9	AC010969	AC010969 Homo sapi
34	35.8	7.2	190790	10	AC102582	AC102582 Mus muscu
35	35.8	7.2	214496	10	AL773520	AL773520 Mouse DNA
36	35.6	7.1	114932	2	AC116344	AC116344 Homo sapi
37	35.6	7.1	145514	9	AC099509	AC099509 Homo sapi
38	35.4	7.1	2000	6	AX655393	AX655393 Sequence
39	35.4	7.1	100975	8	AC007980	AC007980 Arabidops
40	35.4	7.1	111876	8	ATF4110	ATF4110 Arabidops
41	35.4	7.1	192731	9	AC026698	AC026698 Homo sapi
42	35.4	7.1	199536	8	ATCHRIV79	ATCHRIV79 Arabidops
43	35.4	7.1	226345	2	AC005406	AC005406 Homo sapi
44	35.4	7.1	274814	2	AC133059	AC133059 Rattus no
45	35.2	7.0	201	11	BV200564	BV200564 squm20443

ALIGNMENTS

RESULT 1	AF213627	Arabisopsis thaliana (thale cress)	6554 bp	mrna	linear	PLN 01-JUN-2000
LOCUS	AF213627	Arabisopsis thaliana MOM mRNA, complete cds.				
DEFINITION	AF213627	Arabisopsis thaliana MOM mRNA, complete cds.				
ACCESSION	AF213627	Arabisopsis thaliana MOM mRNA, complete cds.				
VERSION	AF213627.1	GI:8132767				
KEYWORDS		Arabisopsis thaliana (thale cress)				
SOURCE		Arabisopsis thaliana				
ORGANISM		Arabisopsis thaliana				
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.				
AUTHORS		Amedeo, P., Habu, Y., Afsar, K., Scheid, O. M. and Paszkowski, J.				
TITLE		Disruption of the plant gene MOM releases transcriptional silencing of methylated genes				
JOURNAL		Nature 405 (6783), 203-206 (2000)				
MEDLINE		20279299				
PUBMED		10821279				
REFERENCE		2 (bases 1 to 6554)				
AUTHORS		Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and Paszkowski, J.				
TITLE		Direct Submission				
JOURNAL		Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland				
FEATURES		Location/Qualifiers				
source		1..6554				
		/organism="Arabidopsis thaliana"				
		/mol_type="mRNA"				
		/cultivar="Zhuerich"				
		/db_xref="taxon:3702"				
		/chromosome="1"				
		/map="between ATEAT1 and phyA"				
		310..6315				
		/function="involved in the maintenance of transcriptional gene silencing"				
		/codon start=1				
		/product="MOM"				
		/protein_id="AAF73380.1"				
		/db_xref="GI:8132768"				
		/translation="MKDKKIGLGTGRTIVTRSLAASIPASVQETPLRRSSRGTPST KVTTPASATRKSLRAPSASVSKSGGIVQKSTPSSLRNRNGKTEVLSQSKGSDN				

CDS

SIRKGDSPDIQRKDSVEESTDKIKPIMARSYRALFRGKLKESALVDASPNBEEL
VVVGGSRRIIPAGNDVQGTDCPPADAGSKLPVDETSLDKGTDFPLKSVTETKIV
LDASPIVETGSDVSSELENLETKLOKGTDCSPANAESKTLPVGETSVLEKEYPQ
KFQDNTDCLPPANAEKSLPVGETSLEKDTDFPLKSTTETGKMWLYASPIVETRDS
VICSPSTNLETKQLKSTGLEDIVLPLKRRDTHAIELDACATVANGDDHVMSSDG
VIPSFGKNDNRKPCNTCKRKQVNGDCNRVSCVIVQPVESDNTQDMKETGP
VTSRYEENGQIOHGKSDPKFYSSVPEYVVPVQLSDVQLEQYQOTLFSKLSLSL
SKIDIGALEEELNVRKTCDFHVPMDASLTKNLEHEILDVEIKASGKLHLDDK
MLTHIKNGKLAVFYQATOTPEGLLGNILEDVFGORFGPKSYEHGIYSSKNGAIN
NFNKSQCCVLLLETRACQTLKLRADAFILFGSLNPSHDVKHVKIKIBSCERT
KIFRLYSVCTVEEKALILARQNRQNAVENLRSLTHALLMGASLFDKLDHFFHS
ETPDSVSFEQIMDVIHFEFSSILSSKGEENEVKLLLEAKAAGQYSDSTLFG
EDHILKLSDESNIPFWSKLLGGKPMKYPSTPQNRKRQYVFGESASPKTGCGN
AKHKKASDDVDPRTVPDPPVDDERKASGDMGMALESFKVITLQSSCSGTGDTL
DGNDAFLGMSHISGIPEDMLASQDMGKIPDESORRLHTVLPKMAKLCOVHLSD
ACTSMVGNFLEVYENHRIYERPAITFQALSWIALALVKQLSHKESLVRANSE
LAFKSRVEVDIYISILSCMSKSLFLEHTQGLQDCFGTNSKQSVSTKLVNLSGAT
VDEKINTKMSNSEDSECEMTEKRSCHYSTATRDIEKTSIGIKKKIKKQVQLVQEH
BEKKMELNMVADKKQKLETSKSAVAIRITCSRTSTQVGLKLDHNYERKFDIEK
ISONECLLEQHEVAKKLAEDACWNRNKSAAKLKVCVPIQSGNKHFGSSGN
ISONAPDVQICNNANVEATYADTNCKASKNVQPEAENTLGTMSGSTQOVHMDVDR
NDETMDSALSREOLTKSOSNEHASITVPEILI PADQREFAALNVHLSQDNCRIT
SAASDEDSVRPEVSQLENLSASPEFSLNREALVTYENRTRSHVGFDTNILDQ
NREDCSDQIEDELAMPVQHLASVETRGAAESDQYQDQICPMPSLSPAGTDPDPAAN
TESENLEAEIEPQSAGSETVETDFAASHQDQVTCPLSSPTGQNPAPENIEGQNI
NTSABPHVAGPDAVESDYAVIDQETMGQAQDASCLPSGSGVTQSDIGANIEQNTTV
AOLPTGSDAVVTGSGPVSQDQADASPMPLSPGNHPTAVNIEGLDNTSVAEPHIS
GSDACHEMTSEBGPQVERSTFANLHEGGEVHSAGVTALVPSILANNTRQIAVQPVQ
IIPPVNDPFLHELEKLRRESNSKTPFEKSKILKAELEKMAEVAQAEFRKFEHVE
AENHTYTKIEKDNKLVIMNKLNAFLSKTDKKVPSGAPGRKIQQAQRAQVSA
LHNYTAPOQLOASFPALVSPALQLOQSSFPAPGAPLQQAQSSFPSSVSRPALL
LNFVCPMPQRPQPLISNIATPSTPATNGLRSPAPHLNSYRPSSTPVATATPTS
SVPPQALITYSAVSIQOQEQOQPSLSSGLSQSNNEVVCLSDDE"

ORIGIN

Query Match 100.0%; Score 500; DB 8; Length 6554;
Best Local Similarity 100.0%; Pred. No. 1.4e-143;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CACAAGCATGAGTTTTCCTCCGGTAATCGTAAATCAAAATCGCTCGTCGAGAAC 60
DB 1 CACAAGCATGAGTTTTCCTCCGGTAATCGTAAATCAAAATCGCTCGTCGAGAAC 60
QY 61 GCGGTGCTAGAGAGGAAAGTACCGTACATAATCTCGGAAACCCCAATTCCTGCTTCTT 120
DB 61 GCGGTGCTAGAGAGGAAAGTACCGTACATAATCTCGGAAACCCCAATTCCTGCTTCTT 120
QY 121 CAAACTCAGTTTCCGAAACCCCAACACCGGAGGATGTCATGGCTGAGAACCACTT 180
DB 121 CAAACTCAGTTTCCGAAACCCCAACACCGGAGGATGTCATGGCTGAGAACCACTT 180
QY 181 AATCGAGATTGCTGGAATTCCTCAAAATTTCCCTCGGTTTCTTTTCACACTCTCG 240
DB 181 AATCGAGATTGCTGGAATTCCTCAAAATTTCCCTCGGTTTCTTTTCACACTCTCG 240
QY 241 AATCGGAATTTCCACCAAGCTCCGTCAAGCGATAGATTCTGCAATTCACACTTTCG 300
DB 241 AATCGGAATTTCCACCAAGCTCCGTCAAGCGATAGATTCTGCAATTCACACTTTCG 300
QY 301 CGAGTGAATATGAAGAAAGTGAAGAGATTGTTTGAACGGGAGAACCAATTCACACGA 360
DB 301 CGAGTGAATATGAAGAAAGTGAAGAGATTGTTTGAACGGGAGAACCAATTCACACGA 360
QY 361 TCCTAGCAGCTTCAATTCCTGCTCAGTTTGAACAGAACCCCTGTTTGAGGAGGTCA 420
DB 361 TCCTAGCAGCTTCAATTCCTGCTCAGTTTGAACAGAACCCCTGTTTGAGGAGGTCA 420
QY 421 AGCCGGGGGACACCATCTAGAAAGGTAATACTCCAGCTTCTGCTACTAGAAAGTCAGAG 480
DB 421 AGCCGGGGGACACCATCTAGAAAGGTAATACTCCAGCTTCTGCTACTAGAAAGTCAGAG 480
QY 481 AGACTGGCTCCCTCAGCTGC 500
DB 481 AGACTGGCTCCCTCAGCTGC 500

Db 481 AGACTGGCTCCCTCAGCTGC 500
RESULT 2
AX063583
LOCUS
DEFINITION
AX063583
ACCESSION
VERSION
AX063583.1 GI:12541307
KEYWORDS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1
AUTHORS
Habu, Y., mittelsten Scheid, O., Amodeo, P. and Paszkowski, J.
TITLE
Gene involved in epigenetic gene silencing
JOURNAL
Patent: WO 0100801-A 2 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
Location/Qualifiers
source
1..6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310..6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
CDS
1..6315
/translation="MKDKEKIGLTGRTTYITRSLAASIPASVEQETFGRLRRSRGTPT
KVITPASATRKSERLAPASVSKSGGIVKNSPTSSLRNRGRTVSLQSSKGSND
SIRKGDSPDIQRKDSVEESTDKIKPIMARSYRALFRGKLKESALVDASPNBEEL
VVVGGSRRIIPAGNDVQGTDCPPADAGSKLPVDETSLDKGTDFPLKSVTETKIV
LDASPIVETGSDVSSELENLETKLOKGTDCSPANAESKTLPVGETSVLEKEYPQ
KFQDNTDCLPPANAEKSLPVGETSLEKDTDFPLKSTTETGKMWLYASPIVETRDS
VICSPSTNLETKQLKSTGLEDIVLPLKRRDTHAIELDACATVANGDDHVMSSDG
VIPSFGKNDNRKPCNTCKRKQVNGDCNRVSCVIVQPVESDNTQDMKETGP
VTSRYEENGQIOHGKSDPKFYSSVPEYVVPVQLSDVQLEQYQOTLFSKLSLSL
SKIDIGALEEELNVRKTCDFHVPMDASLTKNLEHEILDVEIKASGKLHLDDK
MLTHIKNGKLAVFYQATOTPEGLLGNILEDVFGORFGPKSYEHGIYSSKNGAIN
NFNKSQCCVLLLETRACQTLKLRADAFILFGSLNPSHDVKHVKIKIBSCERT
KIFRLYSVCTVEEKALILARQNRQNAVENLRSLTHALLMGASLFDKLDHFFHS
ETPDSVSFEQIMDVIHFEFSSILSSKGEENEVKLLLEAKAAGQYSDSTLFG
EDHILKLSDESNIPFWSKLLGGKPMKYPSTPQNRKRQYVFGESASPKTGCGN
AKHKKASDDVDPRTVPDPPVDDERKASGDMGMALESFKVITLQSSCSGTGDTL
DGNDAFLGMSHISGIPEDMLASQDMGKIPDESORRLHTVLPKMAKLCOVHLSD
ACTSMVGNFLEVYENHRIYERPAITFQALSWIALALVKQLSHKESLVRANSE
LAFKSRVEVDIYISILSCMSKSLFLEHTQGLQDCFGTNSKQSVSTKLVNLSGAT
VDEKINTKMSNSEDSECEMTEKRSCHYSTATRDIEKTSIGIKKKIKKQVQLVQEH
BEKKMELNMVADKKQKLETSKSAVAIRITCSRTSTQVGLKLDHNYERKFDIEK
ISONECLLEQHEVAKKLAEDACWNRNKSAAKLKVCVPIQSGNKHFGSSGN
ISONAPDVQICNNANVEATYADTNCKASKNVQPEAENTLGTMSGSTQOVHMDVDR
NDETMDSALSREOLTKSOSNEHASITVPEILI PADQREFAALNVHLSQDNCRIT
SAASDEDSVRPEVSQLENLSASPEFSLNREALVTYENRTRSHVGFDTNILDQ
NREDCSDQIEDELAMPVQHLASVETRGAAESDQYQDQICPMPSLSPAGTDPDPAAN
TESENLEAEIEPQSAGSETVETDFAASHQDQVTCPLSSPTGQNPAPENIEGQNI
NTSABPHVAGPDAVESDYAVIDQETMGQAQDASCLPSGSGVTQSDIGANIEQNTTV
AOLPTGSDAVVTGSGPVSQDQADASPMPLSPGNHPTAVNIEGLDNTSVAEPHIS
GSDACHEMTSEBGPQVERSTFANLHEGGEVHSAGVTALVPSILANNTRQIAVQPVQ
IIPPVNDPFLHELEKLRRESNSKTPFEKSKILKAELEKMAEVAQAEFRKFEHVE
AENHTYTKIEKDNKLVIMNKLNAFLSKTDKKVPSGAPGRKIQQAQRAQVSA
LHNYTAPOQLOASFPALVSPALQLOQSSFPAPGAPLQQAQSSFPSSVSRPALL
LNFVCPMPQRPQPLISNIATPSTPATNGLRSPAPHLNSYRPSSTPVATATPTS
SVPPQALITYSAVSIQOQEQOQPSLSSGLSQSNNEVVCLSDDE"
ORIGIN
Query Match 100.0%; Score 500; DB 6; Length 6571;
Best Local Similarity 100.0%; Pred. No. 1.4e-143;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CACAAGCATGAGTTTTCCTCCGGTAATCGTAAATCAAAATCGCTCGTCGAGAAC 60

9285. .9808)
/function="involved in the maintenance of transcriptional
gene silencing"
/codon start=1
/product="MOM"
/protein_id="AAF73381.1"
/db_xref="GI:8132770"
/translation="MKDKIKIGLTGRTIYTRSLAASIPASVEQETPGLRRSRGTPT
KVTIPASATRKERLAPSPASVSKSGGIVKNSPTSSLRNRGTEVLSQSKGSDN
STRKGTSPDIORSDVSEEDTKIPIMARSYRALFRGLKESALVDAPNPEELV
VUVGSRRIIPAGNDVOGKTCPPADAGSKRLPYDETSLDKGTGTPPLKATVETEKIV
LDAPIVETGDSVIGSPSENLETKLQDKTDCSPANAESKTLIPVGTSLKXPYQ
KFQDNTDCLPPANAEKRLPVGETSLEKDTDFPLKSTTETGKWLVAIPVETRDDS
VTCPSNTLETKLAVKGTLETDIVLPLKRRDIAEILDACTAVANGDTPDKETGP
VTSREYENGQIQHGKSDPKFVSVPYVWVOLSDVQLEQYCTPLFKSLISLSL
SKIDILGALBETLNSVRKCTDHPYWDASIKOLLTKNLEHILDLVEIKASGKLHLLDK
MUTHIKONGKAVFYFQATQCTPEGLLGNILEDFVQQRGPKSYEHGIVSSKKNAIN
NFKNSQCCVLLLETRACQTKILRADAFILFGSLNSPNSHDVKHVKIKISCSERT
KIFRLYSVCTVEKALILARONKQKAVENLNRLTHALLMWSYLFKLDHPHSS
ETPDSGVSEQIMDGVIIHEFSGILSSKGGEENEVKCLLEAKHAQGYTSDSLFG
EHIKLSDBESPIFWSKLGKGNPMKVPSTPONRKRVOYFEGSEASPKTGDCGN
AKRKKASDDVTPVTPVDDDERKASGDHMGALSPKVTILQSSCKSTGDTL
GNDAPGLKSMGSHISGIFEDMLASQDWKIPDESORRLHTVLKPKAKLCQVHLSD
ACTMGSNVELEYIENHRIYEBPATTFQAFIALSWIAALLVKQLLSHESLVRANSE
LAFKGSRVEDVIYLSLSCMSLFLHTQGLQDFCTNSKQSVSTKLINBSLGSAT
VDEKINTKSMNSDEDECKRCSHYSTATROIEKTIKIKKKYQVOKVQJHEIK
BEKKWLLNMYADKKOKLTSKSVAAVIRITCSRTSTQVGLKLLDHNRYKFDIEIK
SEKNECLSKLEQMDHVAKKLADEACWLNRIKSWAALKVCVPIQSGNNKHPSSGN
ISQNPADVQICNNANVEATYADTNMCAKVNQVPEANLITGWSGSTQVHEMDVVR
NDETMDSALSREQLTQSNEHASITVPEILIPADCBEEFALNVLHSEDCQDKRIT
SAASDEDSRVPVEVQSLENLSASPEFLNREELVTTENRTSHVSGFTDNDILQQ
NREDCSLDQIEIPDELAMPVHLASVETRGAAESDQYQDQICPMPSLACKGDPAAAN
TESENLEERAEPOSAGEVETTPAASHQGVQVTPCLSSPTGNQOPEANIEGONI
NTSAEPHAGPNAVESGVYAVIDQETWGAQACSLPQSGVGTQSDLGANIEQONTTV
AQTPTGSDAVTVGSGFVSDQCADSPMLSPSGNHPDTAVNIEGLDNTSVAEPHIS
GSDACMEISFEGPQVETSTFANLPHFEGVHSAGVTALVPSLLNNGTQIAVQVPVQ
IFPVPNDPFLHEKLERSEANSKKTFFEEKSILKAELERKMAEVQAEFRKFEVE
ASHNTTITKIEKDKNLVINMKLIANAFLSKCTDKKVPSPGACRGKIQQLAQRAAAQV
LHNYTAPOLOASPPAPALVAPLOQSPFAPGAPLOQPAOSFPSSVSRPSALL
LNPVCPMPQBPQPLSNIAPTPSVTPATNPLGRLSPAPHLNSYRSPSSTPVATATPTS
SVPPQALITYSAVSIQQEQQFQOQSLSGLOSNEVVCISDDE"

ORIGIN

Query Match 60.8%; Score 304; DB 8; Length 10329;
Best Local Similarity 100.0%; Pred. No. 1.1e-82;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CACAAGCATGAGTTTTCCTTCGGTAATCGTAAATCAAAATCGCTCGCTCGAGAAC 60
DB 705 CACAAGCATGAGTTTTCCTTCGGTAATCGTAAATCAAAATCGCTCGCTCGAGAAC 764
QY 61 GCGGTGCTAGAGAGGAAGTACCTGACATATCTCGGAACCCCAATTCCTGCTTCCT 120
DB 765 GCGGTGCTAGAGAGGAAGTACCTGACATATCTCGGAACCCCAATTCCTGCTTCCT 824
QY 121 CAAACTCAGTTTTTCCGAAACCCCAACACCGCAGGAGTATGATGGCTCGAAGAACCACTT 180
DB 825 CAAACTCAGTTTTTCCGAAACCCCAACACCGCAGGAGTATGATGGCTCGAAGAACCACTT 884
QY 181 AATCGAAGATTGCTGGAAATTCCTCAAAATTTCCCTCGCGGTTTTTCTTTCACACTCTCGG 240
DB 885 AATCGAAGATTGCTGGAAATTCCTCAAAATTTCCCTCGCGGTTTTTCTTTCACACTCTCGG 944
QY 241 AATCGAAGATTTCACCAAGCTCCGTCAGCGATAGATTCTGACAAATTACACACTTTCGC 300
DB 945 AATCGAAGATTTCACCAAGCTCCGTCAGCGATAGATTCTGACAAATTACACACTTTCGC 1004
QY 301 GCAG 304
DB 1005 GCAG 1008

RESULT 5
AC026875 120965 bp DNA linear PLN 28-JUN-2000
LOCUS Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
DEFINITION 1, complete sequence.
ACCESSION AC026875.4 GI:7651711
VERSION AC026875
SOURCE HTG.
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 120965)
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C.,
Khan, S., Kim, C., Altafi, H., Bei, Q., Chin, C., Chiou, J., Choi, E.,
Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T.,
Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,
Mukharaky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J.,
Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G.,
Federspiel, N.A., Theologis, A. and Ecker, J.R.
Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
1
1 Unpublished
2 (bases 1 to 120965)
Ecker, J.R.
Direct Submission
Submitted (25-MAR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 120965)
Ecker, J.R.
Direct Submission
Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 120965)
Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q.,
Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C.,
Chiou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N.,
Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,
Liu, S., Mukharaky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,
Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M.,
Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.
Direct Submission
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Apr 27, 2000 this sequence version replaced gi:7543638.
COMMENT Location/Qualifiers
FEATURES
source
1. 120965
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T6D22"
join(410..559,596..758,828..865,1681..1731,1903..2253,
2562..2999,3096..3146,3240..3322,3493..3659,3781..3819,
3992..4021,4069..4246,4597..4703,4824..4952,5058..5122,
5219..5394,5549..5709,5812..5843,6103..6266,6382..6461,
6588..6723,6881..7170,7235..7264,7463..7570,7663..7760,
7939..7999,8364..8440,8552..8623,8702..8778,8898..8978,
9337..9439,9593..9751,9883..10281)
/note="similar to AT.124-9 gene product gb|AA07881.1;
similar to EST emb|Z35021.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.1"
/protein_id="AAF79821.1"
/db_xref="GI:8778822"
/translation="MIGHNFCGRNRLDSLFLPTTKNSLSQTVIVAPKISFVNRCS
TQVTGKICVTKPSSSSFPNQADKCGKRVVAASRVCVGLLRILNKLQNFKPRD
LYNKVILTQIGDGGFMLVEPDGFSFDLASKAFIDFCLLRWSPPHNTSTNSVVVPPV

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrgeorgie, E., Geer, K., Gill, R., Grady, W., Guerra, T., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, N., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., Liu, W., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louissegh, M., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Reigh, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 221873)
Worley, K.C.

Direct Submission
Submitted (23-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221873)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 9, 2003 this sequence version replaced gi:24942204.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GAVN
Center clone name: CH230-4H24
----- Summary Statistics
Assembly program: Atlas
Consensus quality: 210371 bases at least Q40
Consensus quality: 213256 bases at least Q30
Consensus quality: 214662 bases at least Q20
Estimated insert size: 219784; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 220351: contig of 220351 bp in length
* 220352 220451: gap of unknown length
* 220452 221873: contig of 1422 bp in length.

FEATURES
source
1..221873
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4H24"
214451..215686
/note="wgs contig"
217841..220351
/note="wgs_contig"
ORIGIN
Query Match 7.6%; Score 38.2; DB 2; Length 221873;
Best Local Similarity 55.7%; Pred. No. 6.1;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 25 GTAATCGTAAATCAAAAATCGCTCGTCGAGAAACGCGGTCTAGAGAGAGAAAGTAC 84
|||||
Db 34262 GTAATGTATATATACAAAAGACTTCTTAGAACACAGAAATTAAGAAGCAGGCC 34203
QY 85 CGTACATATCTCGGAACCCCAATTCGTCTCTTCTTCAAACTCAGTTTTCCGAAACCCCA 144
Db 34202 TCTGCATCCCAAGTAACACACACAGTCGCTGTTTCCACACAGACCCCAAGTACC 34143
QY 145 AACACCGCGAG 155
Db 34142 AACCTGCTGAG 34132
RESULT 8
AC094977
LOCUS
Rattus norvegicus clone CH230-6K1, WORKING DRAFT SEQUENCE, 4
DEFINITION
unordered pieces.
ACCESSION
AC094977
VERSION
AC094977.5 GI:30467201
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 225918)
AUTHORS
Muzny, D., Maric, Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

RESULT 9
AC131437
LOCUS
DEFINITION Rattus norvegicus clone CH230-7A7, *** SEQUENCING IN PROGRESS ***,
2 unordered pieces.
AC131437 234966 bp DNA linear HTG 09-MAY-2003
VERSION AC131437.4 GI:30467830
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 234966)
Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsebrook,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Daviia,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgieff,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollands,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,X., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 234966)
Rat Genome Sequencing Consortium.
TITLE
AUTHORS
JOURNAL
Submitted (22-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 234966)
Rat Genome Sequencing Consortium.
TITLE
AUTHORS
JOURNAL

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24941153.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCGT
Center clone name: CH230-7A7
----- Summary Statistics
Assembly program: Atlas
Consensus quality: 210598 bases at least Q40
Consensus quality: 213297 bases at least Q30
Consensus quality: 215090 bases at least Q20
Estimated insert size: 221930; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 233801: contig of 233801 bp in length
* 233802 233901: gap of unknown length
* 233902 234966: contig of 1065 bp in length.
FEATURES
source
1. .234966
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-7A7"
misc_feature
1. .1725
/notes="wgs contig"
misc_feature
5145. .7136
/notes="wgs contig"
misc_feature
232638. .233801
/notes="wgs contig"
ORIGIN
Query Match 7.6%; Score 38.2; DB 2; Length 234966;
Best Local Similarity 55.7%; Pred. No. 6.1;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 25 GTAATCGTAATAATCAAAATCGCTCGAGAAACGCCGGTCTAGACAGAAAGTAC 84
DB 224134 GTAATGTGATATAACAAAGAACTTCTCTAGGAACAACAGAAATTAAAAAGCAAGGCC 224193
QY 85 CGTACATAATCTCTGCAACCCCAATTCCTGCTCTCTCTTCAAACTCAGTTTCCGAAACCCCA 144
DB 224194 TCTGATATCCCAAGTAACCAACACACAGTCGCTGTTTCCCAACAGACCCCAAGTACC 224253
QY 145 AACACCCGAG 155

```

Db      224254 AACCTGCTGAG 224264      |||
                                           |||
RESULT 10
LOCUS   AC006233
DEFINITION Arabidopsis thaliana chromosome 2 clone F12K2 map B68, complete
sequence.
ACCESSION AC006233
VERSION   AC006233.4
KEYWORDS  HTG.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 88383)
AUTHORS  Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
          Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
          Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
          Nierman,W.C., Fraser,C.M. and Venter,J.C.
          Unpublished
JOURNAL  2 (bases 1 to 88383)
REFERENCE 1. 88383
AUTHORS  Lin,X.
TITLE    Direct Submission
JOURNAL  Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
          Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 88383)
AUTHORS  Town,C.D. and Kaul,S.
TITLE    Direct Submission
JOURNAL  Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
          Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT  On Apr 18, 2002 this sequence version replaced gi:6598539.
FEATURES             Location/Qualifiers
     source           1..88383
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="B68"
                     /clone="F12K2"
     misc_feature     1..31423
                     /notes="overlap with BAC clone F10A12 (AC006232:1..31422)."
     repeat_region    293..312
                     /rpt_family="A)n"
     gene             741..1780
                     /genes="At2g27410"
                     /notes="synonym: F12K2.1; predicted by genefinder"
     mRNA             join(<741..1079,1190..>1780)
     CDS              join(741..1079,1190..1780)
                     /genes="At2g27410"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD41990.1"
                     /db_xref="GI:5306258"
                     /translation="MAKPGNRNKNVKKHDDLIVLSMRISDTELLLEIVQRVENNYA
                     NHTESQTTSFGKQVDPNIGDAVPIKRGPCRSRSSSFTKTTICTRALEQQRSLTS
                     VFVLKTLIEEPIQTTPPELLANMRREENGYNPKLISTRQLYKTDLKKTEARLSV
                     PFQVTPDPLFDETRIIEHNAKMRDNGVPVNFVDPELNKHVLELRKWKQKGNWYI
                     VFVKGWNLDAKTLFKEDDVYPLMSFRSGTGKLCFALTTPKNSGRGNSLPGDGAST
                     SGEQGVPLPIPPARYSSNSGQCSGESSSSS"
                     5102..5375
                     /genes="At2g27395"
                     /notes="synonym: F12K2.20"
     mRNA             join(<5102..5253,5267..>5375)
                     /genes="At2g27395"
     CDS              join(5102..5253,5267..5375)
                     /genes="At2g27395"
                     /codon_start=1
                     /product="putative cysteine proteinase"

```

```

/protein_id="AAM15194.1"
/db_xref="GI:20197673"
/translation="MNNEALVQAVSOHPVSMIGETGAARFHYSGGIFNGRCETDFA
SCGYNCCREGTKYMWVKNVSGENGFMRIKRDVDAQGMYYN"
5538..5942
/genes="At2g27390"
/notes="synonym: F12K2.3; predicted by genefinder"
<5538..>5942
/genes="At2g27390"
5538..5942
/genes="At2g27390"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD41991.1"
/db_xref="GI:5306259"
/translation="MKMNTKEYSIRTPSLGNLPQTGTCCPPPLVFPPLPLSPSPS
PPSPSPRLPPFPALFPPEPLPPPELPPPLFPPLPRLPPLPLPPPEPPE
PPPPPPPEPPPPASCRLTKSPENGIVSTVK"
complement(5640..5710)
/rpt_family="TGG)n"
complement(5760..5867)
/rpt_family="GGA)n"
8239..8260
/rpt_family="AT_rich"
8352..8385
/rpt_family="AT_rich"
complement(10730..13015)
/genes="At2g27380"
/notes="synonym: F12K2.4; predicted by genefinder and
geneScan"
complement(<10730..>13015)
/genes="At2g27380"
complement(10730..13015)
/genes="At2g27380"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD41992.1"
/db_xref="GI:5306260"
/translation="MRVPLIDFLRLVILSLGSAAADATVKQFNKYETDSGHAH
PPPIYGAPSVYTPPPPIYSPPIYSPPIYSPPIYSPPIYSPPIYSPPIYSPPIY
PIQKPTPTSPPIYSPPIYSPPIYSPPIYSPPIYSPPIYSPPIYSPPIYSPPIY
PTISPIKPPVHKPPTTSPPIKPPVHKPPTTSPPIKPPVHKPPTTSPPIKPPV
PPVHKPPTTSPPIKPPVHKPPTTSPPIKPPVHKPPTTSPPIKPPVHKPPTTSP
PTIYSPVKKPPTTSPPIYSPVKKPPTTSPPIYSPVKKPPTTSPPIYSPVKKPPT
PIKPPVKKPPTTSPPIYSPVKKPPTTSPPIYSPVKKPPTTSPPIYSPVKKPPTI
KPTTYSBPPIKPPLOKPTTSPPIKPPVKKPPTTSPPIYSPVKKPPTTSPPIYSP
PEVKPPVHKPPTTSPPIKPPVKKPPTTSPPIYSPVKKPPTTSPPIYSPVKKPPT
QKPTTYSBPPIKPPVKKPPTTSPPIKPPVKKPPTTSPPIKPPVHKPPTTSPPI
SPPIKPPVHKPPTTSPPIKPPVKKPPTTSPPIKPPVKKPPTTSPPIKPPVHKP
PVHKPPTTSPPIKPPVKKPPTTSPPIKPPVKKPPTTSPPIKPPVKKPPTTSPPI
PTIYSPVKKPPTTSPPIYSPVKKPPTTSPPIKPPVKKPPTTSPPIYSPVKKPPT
QGGYGTTPPPYAYLSHPIDIRN"
complement(15102..15752)
/genes="At2g27375"
/notes="synonym: F12K2.2"
complement(join(<15102..15341,15576..>15752))
/genes="At2g27375"
complement(join(15102..15341,15576..15752))
/genes="At2g27375"
/codon_start=1
/product="putative reverse transcriptase"
/protein_id="AAM15193.1"
/db_xref="GI:20197671"
/translation="MEVWFFCQEVVETREHLFFSRPYSTQVMSALTDQMLGWFTRW
BSLIPLIWASTTPOVSOASHLIGISKLLASDIQQLVKALNGAHPNPKRGILHILHI
SLTFAEISGFVYKRNKLNKEDALSKALNSISSGPI"
16521..17976
/genes="At2g27370"
/notes="synonym: F12K2.5"
join(<16521..16843,17190..17313,17580..17976)
/genes="At2g27370"
join(16521..16843,17190..17313,17580..17998)
/genes="At2g27370"

```

```
/codon_start=1
/product="unknown protein"
/protein_id="AAD41993.1"
/db_xref="GI:5306261"
/translations="MDIEKAGSRREBEPIVQPKDKGKAHVAPPMVNYNRIMDK
HKOKRSPAGKRGVAFDFVRLIAITAMAAAKMATEETLPETLOFQADYT
DLPTMSFVIVNSIGVGLTSLPFSVICILRLPVLAVPRLFLILCDTVMGLTMAAS
ASAAIVYLHNGSSNWLPVCOQFGDGTSGAVVASFIAATLLMLFLVLSAFALK
RTT"
repeat_region 17319..17357
/rpt_family="AT_rich"
complement(18113..18141)
/rpt_family="AT_rich"
18391..18413
/rpt_family="(CAATA)n"
complement(18521..20193)
/gene="At2g27360"
/notes="synonym: F12K2.6"
complement(join<18521..18801,18878..19154,19249..19397,
19496..19714,19935..20193))
/gene="At2g27360"
complement(join(18521..18801,18878..19154,19249..19397,
19496..19714,19935..20193))
/gene="At2g27360"
/codon_start=1
/product="putative lipase"
/protein_id="AAD41994.1"
/db_xref="GI:5306262"
/translations="MASOCHMLSPFISTFLITVTSOTCRNFKSIISFGDSITDT
GNLIGSSNDLPESAPPYGTEFHHPSRFSDDGLIIDTAEFLGIPHPFFYGSK
NMFKGVNFAVGATDACSVELEKGTCSQNSISGNQLSKFSLSSGSSSPD
CRDMENAFILGEIGNDYNDPFLDRKNIIEVKELPVLITTSISALBVDMGART
FLVGNFPIGDSVAYLTLYETNKEEYFLTGLTDLNDFSVYHQDLAEKRLNL
YHVNIIYGDYNTLLRLMOEPSKFLMDRPLPACCGLGNGYNTFSIKGSGVYVC
SPSKYVNDGIHMTAAVKWISGVLTCGYAIPPPFNWSCLSKIKNNESLHTQYSLM
NS"
20776..20796
/rpt_family="AT_rich"
complement(20914..20946)
/rpt_family="AT_rich"
22195..26847
/gene="At2g27350"
/notes="synonym: F12K2.7; similar to ovarian tumor protein;
supported by cDNA: gi.16604333.gb_AY058065.1"
join(22195..22369,24044..24662,24742..24903,25419..25496,
25583..25661,25752..25897,26025..26090,26174..26345,
26432..26847)
/gene="At2g27350"
join(24083..24662,24742..24903,25419..25496,25583..25661,
25752..25897,26025..26090,26174..26345,26432..26646)
/gene="At2g27350"
/codon_start=1
/product="expressed protein"
/protein_id="AAD41995.2"
/db_xref="GI:20197675"
/translations="MTRILVQRSGSSNSRPPSSSSSSSGSETQINNIPVPPVT
IDEEITDEQEEVTVVEKADKAVDSDPADREDDGLVAENVHVSQEGIDC
DSPVSGSSSPVPVPPKPSPTVNGSNRSLVSGFALRIGPTRRAAGPSLSVSS
RSPSTGSHPSRSHSENGVSSDEHPCYVPSHPGSLEREHQPFAIRYSKPEI
RMLEDGNCILFRAVDQVVGDSAYDLARCMQMDINEQRDHFQSQITGFTSYLKRK"
Query Match 7.5%; Score 37.6; DB 8; Length 88383;
Best Local Similarity 55.3%; Pred. No. 7.9;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 199 AATCTCAATTTTCCTCGCGTTCCTTTCTTTCACACTCTCGAATCGGAATTCACCA 258
DB 18285 AATTTGTTTATATGTCCTTTTATTATCTCTATCCAACTTTGAAATATCTTAATCCAACTA 18344
QY 259 AGCTCCGTCAGCGATAGATTCTGACAAATTTACACATTTTCGCCAGTGAATAGAGAA 318
DB 18345 AGATACCATGATTTAGCAACTACCACCTTTTACCACCTTTTCCCAACATATATACAA 18404
```

QY 319 GATGAAAAGATT 330
Db 18405 TACAATACATT 18416

RESULT 11
AC006232/c
LOCUS Arabidopsis thaliana chromosome 2 clone F10A12 map B68, complete
DEFINITION AC006232
AC006232
VERSION AC006232.4 GI:20197647
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 90341)
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
Nierman,W.C., Fraser,C.M. and Venter,J.C.
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 90341)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 90341)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdowmetigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598338.
FEATURES
Location/Qualifiers
1..90341
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="B68"
/clone="F10A12"
complement(1..31422)
/note="overlap with BAC clone F12K2 (AC006233.1..31423)."
complement(31..1042)
/gene="At2g27330"
/note="synonym: F10A12.1; supported by full length cDNA:
Ceres:18569"
complement(join<31..443,747..853,1020..1042))
/gene="At2g27330"
complement(join(348..443,747..842))
/gene="At2g27330"
/codon_start=1
/product="putative RNA-binding glycine rich protein
(RGP-2)"
/protein_id="AAM15185.1"
/db_xref="GI:20197660"
/translation="MDKIRCPKGFAYVTFSSKEAKALLELNALQVDRVILDTT
KAAKNPPDKPNHAGGS"
misc_feature 1312..1502
/note="molecular marker ngal126"
repeat_region 1449..1481
/rpt_family="(GN)n"
2447..4118
/gene="At2g27340"
/note="synonym: F10A12.2"
join<2447..2556,2782..2852,3021..3148,3473..3530,
3709..3925,4031..4118)
/gene="At2g27340"
join(2447..2556,2782..2852,3021..3148,3473..3530,
3709..3925,4031..4118)
CDS


```

QY      319 GATGAAAGATT 330
Db      13018 TACAATACATT 13007

RESULT 12
LOCUS   HSJ322G13
DEFINITION
HSJ322G13      112180 bp      DNA      linear      PRI 19-MAR-2001
Human DNA sequence from clone RP3-322G13 on chromosome
20p11.21-12.3 Contains the gene for NTF2-related export protein
(NXT1), a gene for zinc finger protein FLJ211794, two putative novel
genes, the gene for beta-soluble NSF attachment protein
(SNAP-beta), a novel gene for a protein similar to cystatin,
another novel gene for a protein similar to cystatin 8 (CST8),
ESTs, STSS, GSSs and six CpG islands, complete sequence.

ACCESSION
VERSION      AL096677
KEYWORDS     AL096677.21 GI:10862836
SOURCE      HTG; CpG island; CST8; cystatin; export protein; FLJ211794; NSF
            attachment protein; NTF2; NXT1; SNAP; zinc finger protein.
ORGANISM     Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 112180)
REFERENCE    Lovell, J.
AUTHORS      Direct Submission
TITLE        Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL      CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonequery@sanger.ac.uk
COMMENT      On Oct 17, 2000 this sequence version replaced gi:9795218.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; Swi: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence is the entire insert of clone RP3-322G13 The true
            left end of clone RP3-333B15 is at 28739 in this sequence. The true
            right end of clone RP3-851M4 is at 70589 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP3-322G13 is from
            the library RPCI-3 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: PCVPAC2.
FEATURES             Location/Qualifiers
     source           1..112180
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosomes="20"
                     /map="p11.21-12.3"
                     /clone="RP3-322G13"
                     /clone_lib="RPCI-3"
     repeat_region    79..144
                     /note="LTR16C repeat: matches 321..387 of consensus"
     repeat_region    790..1085
                     /note="AluSg repeat: matches 1..300 of consensus"
repeat_region        1247..1357
                     /note="MLT1H repeat: matches 384..532 of consensus"
repeat_region        1748..1831
                     /note="21 copies 4 mer gtgt 69% conserved"
repeat_region        1757..1826
                     /note="5 copies 14 mer 75% conserved"
gene                 complement(join(2155..2811,6544..6700))
                     /genes="dj322G13.7"
mRNA                 complement(join(<2155..2811,6544..6700))
                     /genes="dj322G13.7"
                     /product="dj322G13.7 [putative novel transcript]"
                     /evidence=match: ESTs; Em:BE91551"
                     /evidence=not_experimental
repeat_region        2488..2527
                     /note="MIR repeat: matches 108..147 of consensus"
repeat_region        3133..3309
                     /note="MIR repeat: matches 76..254 of consensus"
repeat_region        3727..4165
                     /note="LIR2 repeat: matches 1..449 of consensus"
repeat_region        4279..4345
                     /note="L2 repeat: matches 2679..2745 of consensus"
misc_feature         5214..6623
                     /note="CpG island"
                     /evidence=not_experimental
gene                 join(6067..6275,9195..9991)
                     /genes="dj322G13.1"
mRNA                 join(6067..6275,9195..9991)
                     /product="dj322G13.1 (NTF2-related export protein NXT1)"
                     /note="match: CDNAS; Em:AF158957"
                     /match: ESTs; Em:R35341 Em:AA223265 Em:AW249178"
                     /evidence=not_experimental
repeat_region        7047..7168
                     /note="MIR repeat: matches 99..220 of consensus"
repeat_region        7169..7507
                     /note="MER2 repeat: matches 1..341 of consensus"
repeat_region        7508..7523
                     /note="MIR repeat: matches 220..235 of consensus"
repeat_region        8454..8548
                     /note="MER2 repeat: matches 9..94 of consensus"
repeat_region        8552..8841
                     /note="AluSx repeat: matches 1..292 of consensus"
                     /genes="dj322G13.1"
CDS                 9256..9678
                     /note="match: not_experimental
                     /product="dj322G13.1 (NTF2-related export protein NXT1)"
                     /protein_id="CAC03437.2"
                     /db_xref="GI:10862837"
                     /db_xref="GOA:Q9UKK6"
                     /translation="VAVDFKYYVDQACRAAEFVNYYTTHDKRRLLSRLYNGTAT
                     LKNGNAVSGQSLSEFFLPISQFISVDCQPVHDEATPSQTTLVWVIGSVKXPE
                     GVNRDQNFQFILTAQASPSNTVSKIASDCFRQDWAS"
                     complement(9828..9960)
misc_feature         /note="match: STS; Em:G42938"
repeat_region        10134..10209
                     /note="MER5B repeat: matches 3..82 of consensus"
repeat_region        10219..10270
                     /note="13 copies 4 mer gtgt 96% conserved"
repeat_region        10220..10271
                     /note="26 copies 2 mer tg 98% conserved"
repeat_region        10223..10278
                     /note="4 copies 14 mer 92% conserved"
repeat_region        10299..10431
                     /note="AluSg/x repeat: matches 128..259 of consensus"
repeat_region        10445..10487
                     /note="MER5B repeat: matches 34..76 of consensus"
repeat_region        11048..11269
                     /note="MLT1J repeat: matches 125..364 of consensus"
                     complement(11808..13219)
gene

```

```
/gene="dj322G13.5"
complement(join(<11808..12183,12265..12351,13169..>13219))
/gene="dj322G13.5"
/product="dj322G13.5.1 (putative novel transcript, isoform
1)"
/note="match: ESTs: Em:AI693606 Em:AI984374"
/evidence=not_experimental
complement(join(<11808..12039,12265..12351,12935..>13069))
/gene="dj322G13.5"
/product="dj322G13.5.2 (putative novel transcript, isoform
2)"
/note="match: ESTs: Em:AW170001"
/evidence=not_experimental
12286..12878
/note="match: STS: Em:HS838D8S"
12555..12780
/note="AluSg/x repeat: matches 80..305 of consensus"
12988..13655
/note="CpG island"
/evidence=not_experimental
14120..14849
/note="MER49 repeat: matches 184..916 of consensus"
14936..15089
/note="AluSx repeat: matches 138..282 of consensus"
15090..15383
/note="AluSx repeat: matches 1..294 of consensus"
15384..15518
/note="AluSx repeat: matches 1..138 of consensus"
15657..15698
/note="L2 repeat: matches 2709..2750 of consensus"
15714..16375
/note="MER67C repeat: matches 1..710 of consensus"
16723..17993
/note="CpG island"
/evidence=not_experimental
18803..18837
/note="MIR repeat: matches 126..160 of consensus"
complement(18817..19228)
/note="match: GSS: Em:AQ203877"
complement(18864..19228)
/note="match: GSS: Em:AQ145317"
18914..18973
/note="MIR repeat: matches 74..140 of consensus"
19249..19878
/note="match: GSS: Em:AQ895221"
19598..28277
/gene="dj322G13.2"
join(19598..20961,22217..22311,23976..24143,24798..24955,
25305..28277)
/gene="dj322G13.2"
/product="dj322G13.2.3 (zinc finger protein FLJ21794,
isoform 3)"
/note="match: CDNAS: Em:AL110217 Em:AF167320 Em:X78927
Em:AJ224763 Em:AO80123 Em:AF161544 Em:Z67747 Em:AB037770
Em:X60155 Em:X82125 Em:U66561 Em:AF027146 Em:U36315
Em:M36514 Em:X12592 Em:X64413 Em:AF022158 Em:L75847
Em:D10632 Em:X06021 Em:X89264 Em:X78933 Em:L15309
Em:X52533 Em:S67970 Em:AK025447
match: ESTs: Em:AA417107 Em:R68950 Em:AI023363 Em:AI217180
Em:AW291034 Em:AA417012 Em:AA701588 Em:AI177089
Em:AI145986 Em:AW079910 Em:AI961159 Em:AW019926
Em:AA035213 Em:AW082641 Em:AW594742 Em:AW531679
Query Match 7.5%; Score 37.4; DB 9; Length 112180;
Best Local Similarity 53.8%; Pred. No. 9.5;
Matches 77; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 166 CCTGAGAGACCACTTAATCGAGATTGTGCTGGAATTCCTCAAAATTTTCCTCGCTTTT 225
Db 74611 CTTTAAAGGATTATAAGCTCAAGATTTGGCTAGTTCTCTTAAAAACATCGAAACCTCTTT 74670
QY 226 CTTTCACATCTCGGAATCGGAATTTCCACCAAGCTTCGTCAGCGATAGATTCTGCACA 285
|||||
```

```
Db 74671 CCCAAAAGAGGTGTTCTCTGAGACAAGCACCAAGTGCTGTCAAGCTTTTCTCTGCCA 74730
QY 286 ATTACACACTTTTCGGCGCACTGAA 308
|||||
Db 74731 GTTACAGACTCTCCACACAGGAA 74753

AC008995 39329 bp DNA linear PRI 15-JUL-2000
Homo sapiens chromosome 19 clone LLNLR-238B11, complete sequence.
AC008995
ACCESSION AC008995
VERSION HTG.
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39329)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 39329)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 15, 2000 this sequence version replaced gi:7711391.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES
Location/Qualifiers
1..39329
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLR-238B11"
ORIGIN
Query Match 7.4%; Score 37.2; DB 9; Length 39329;
Best Local Similarity 65.9%; Pred. No. 9.1;
Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 308 ATATGAAGAAGATGAAAGATTGTTTGACGGGAGAACCACTTTACACCAATCCTCTAG 367
Db 28399 ATTACATAGGTTCAAAAGATTGTTGACCGGGTGTTACCATTTACATAAACCCTGCAAAA 28458
QY 368 CAGCTTCAATTCCTGCCTCAGT 389
Db 28459 AAATTGCTCTCCACCCCAAT 28480

AC020700 181532 bp DNA linear HTG 07-JUL-2000
Homo sapiens chromosome 19 clone RP11-157B13, WORKING DRAFT
AC020700
ACCESSION AC020700
VERSION AC020700.4 GI:8569830
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RESULT 14
AC020700/c
LOCUS AC020700
DEFINITION Homo sapiens chromosome 19 clone RP11-157B13, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION AC020700
VERSION AC020700.4 GI:8569830
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```


REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 181532)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 181532)
Waterston,R.H.
Direct Submission
Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7230911.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----
Center project name: H NH0157B13

----- Summary Statistics -----

Sequencing vector: M13; 100%

Chemistry: Dye-terminator; 0%

Chemistry: Dye-terminator; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 171574 bases at least Q40

Consensus quality: 174349 bases at least Q30

Consensus quality: 175918 bases at least Q20

Insert size: 175000; agarose-fp

Insert size: 179432; sum-of-contigs

Quality coverage: 5.00 in Q20 bases; agarose-fp

Quality coverage: 4.93 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 1478: contig of 1478 bp in length
* 1479
* 1578: gap of unknown length
* 1579
* 3027: contig of 1449 bp in length
* 3028
* 3127: gap of unknown length
* 3128
* 4253: contig of 1126 bp in length
* 4254
* 4353: gap of unknown length
* 4354
* 5930: contig of 1577 bp in length
* 5931
* 6030: gap of unknown length
* 6031
* 8127: contig of 2097 bp in length
* 8128
* 10325: gap of unknown length
* 10326
* 10425: gap of unknown length
* 10426
* 12077: contig of 1652 bp in length
* 12078
* 12177: gap of unknown length
* 12178
* 15635: contig of 3458 bp in length
* 15636
* 15736: gap of unknown length
* 15737
* 19769: contig of 4034 bp in length
* 19770
* 19869: gap of unknown length
* 19870
* 25357: contig of 5487 bp in length
* 25358
* 25457: gap of unknown length
* 28614
* 28713: contig of 3157 bp in length
* 28714
* 32798: contig of 4085 bp in length
* 32799
* 32898: gap of unknown length
* 32899
* 40209: contig of 7311 bp in length
* 40210
* 40309: gap of unknown length
* 49179: contig of 8870 bp in length
* 49180
* 49280: gap of unknown length
* 54995: contig of 5716 bp in length
* 54996
* 55095: gap of unknown length
* 55096
* 66240: contig of 11145 bp in length
* 66340: gap of unknown length
* 66341
* 76524: contig of 10184 bp in length

* 76525
* 76625
* 88455: gap of 11831 bp in length
* 88456
* 99745: gap of 11190 bp in length
* 99746
* 99845: gap of unknown length
* 99846
* 117030: contig of 17185 bp in length
* 117031
* 117130: gap of unknown length
* 117131
* 146072: contig of 28942 bp in length
* 146073
* 146172: gap of unknown length
* 146173
* 181532: contig of 35360 bp in length.

FEATURES

source

Location/Qualifiers

1..181532

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="19"

/clone="RP11-157B13"

1..1478

/note="assembly_name:Contig7"

1579..3027

/note="assembly_name:Contig9"

3128..4253

/note="assembly_name:Contig10"

4354..5930

/note="assembly_name:Contig11"

6031..8127

/note="assembly_name:Contig12"

8228..10325

/note="assembly_name:Contig13"

10426..12077

/note="assembly_name:Contig14"

12178..15635

/note="assembly_name:Contig15"

15736..19769

/note="assembly_name:Contig16"

19870..25356

/note="assembly_name:Contig17"

25457..28613

/note="assembly_name:Contig18"

28714..32798

/note="assembly_name:Contig19"

32899..40209

/note="assembly_name:Contig20"

40310..49179

/note="assembly_name:Contig21"

49280..54995

/note="assembly_name:Contig22"

55096..66240

/note="assembly_name:Contig23"

66341..76524

/note="assembly_name:Contig24"

76625..88455

/note="assembly_name:Contig25"

vector_side:left

88556..99745

/note="assembly_name:Contig26"

99846..117030

/note="assembly_name:Contig27"

117131..146072

/note="assembly_name:Contig28"

146173..181532

/note="assembly_name:Contig29"

ORIGIN

Query Match 7.4%; Score 37.2; DB 2; Length 181532;

Best Local Similarity 65.9%; Pred. No. 12;

Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 308 ATATGAAGAAGATGAAGATTGGTTGACGGGGAGNACCATTTACACCATCCCTAG 367

Db 120599 ATTACATAGGGTTCAAAAGATTGGTTGACGGGGTGTACCATTTACATACCTGCAAAA 120540

```

Qy 368 CAGCTTCAATTCTCGCTCAGT 389
Db 120539 AAACCTTGCCTTCCCAACCAAT 120518

RESULT 15
AC073539/c
LOCUS AC073539 184213 bp DNA linear PRI 30-NOV-2000
DEFINITION Homo sapiens chromosome 19 clone RP11-157B13, complete sequence.
ACCESSION AC073539
VERSION AC073539.3 GI:11465052
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184213)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
REFERENCE 2 (bases 1 to 184213)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 184213)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 30, 2000 this sequence version replaced gi:9964807.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
STS Content:
SHGC-9450 GI4419.
FEATURES
source
1..184213
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="19"
/clone="RP11-157B13"

ORIGIN
Query Match 7.4%; Score 37.2; DB 9; Length 184213;
Best Local Similarity 65.9%; Pred. No.12;
Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 308 ATATGAGAAGATGAAGATTGGTTACCGGGGAGACCAATTACACCAAGATCCCTAG 367
Db 25306 ATTACATAGGGTCAAAAGATTGGTTGACCGGGGTGACCAATTACATAACCTGCAAAA 25247

Qy 368 CAGCTTCAATTCTCGCTCAGT 389
Db 25246 AAACCTTGCCTTCCCAACCAAT 25225

Search completed: July 12, 2005, 05:59:39
Job time : 2536.46 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 381.991 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929C-2_COPY_1_500

Perfect score: 500
Sequence: 1 cacaagcatgagttttctc.....agactggctcccaactgc 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	6571	4	AAA89354 Arabidops
2	304	60.8	10329	4	AAA89385 Arabidops
3	304	60.8	10329	4	AAA89353 Arabidops
4	36.4	7.3	30277	11	ACN44648 Mouse gen
5	36.4	7.3	134738	11	ACN44182 Human gen
6	35.4	7.1	2000	8	ADA71938 Rice gene
7	34.2	6.8	1497	10	ADB31378 testoster
8	34	6.8	2339	4	AAK73030 Human imm
9	34	6.8	2339	4	AAK73028 Human imm
10	33.8	6.8	861	2	AAK34746 DNA encod
11	33.8	6.8	861	3	AAC68703 Ehrlichia
12	33.8	6.8	861	6	ABS63279 DNA encod
13	33.8	6.8	4683	3	AAV07179 Ehrlichia
14	33.8	6.8	4683	3	AAC68716 Ehrlichia
15	33.8	6.8	4683	4	AAK07578 DNA encod
16	33.8	6.8	5383	4	ABL13734 Drosophil
17	33.8	6.8	13909	4	ABL20392 Drosophil
18	33.8	6.8	1973	4	ABL29345 Drosophil
19	33.4	6.7	5289	4	ABL29344 Drosophil
20	33.4	6.7	5289	4	ABL29344 Drosophil

21	33.2	6.6	504	4	AAI85190 Human pol
22	33.2	6.6	2801	2	AAV40749 C. felis
23	33.2	6.6	2801	2	AAV40750 C. felis
24	33.2	6.6	2801	4	AAV40745 C. felis
25	33.2	6.6	2836	2	AAV40745 C. felis
26	33.2	6.6	2836	2	AAV40744 C. felis
27	33.2	6.6	2836	4	AAV40744 C. felis
28	33.2	6.6	54355	9	ADA02753 Ctenoceph
29	33.2	6.6	54355	10	ADB72491 Mouse Mor
30	33.2	6.6	54355	10	ADC85233 Mouse Mor
31	33.2	6.6	54355	12	ADM74348 Murine ca
32	33	6.6	806	6	ABQ29272 Oligonuc
33	33	6.6	806	6	ABQ29273 Oligonuc
34	33	6.6	216215	10	ADF69167 Human MP5
35	32.8	6.6	289	2	AAT16521 Mutant sa
36	32.8	6.6	4766	10	ADI40400 Human pur
37	32.6	6.5	478	9	ACH25425 Human adu
38	32.6	6.5	2022	12	ADL15366 B thuring
39	32.6	6.5	2022	12	ADL15368 B thuring
40	32.6	6.5	6793	4	ABL50539 Barley M1
41	32.6	6.5	7153	4	ABL50540 Barley M1
42	32.6	6.5	190000	10	ADL13752 Osteoarth
43	32.4	6.5	880	13	ADQ28759 Green flu
44	32.4	6.5	1071	6	ABZ13233 Arabidops
45	32.2	6.4	2000	6	ABZ16484 Arabidops

ALIGNMENTS

RESULT 1
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
AC AAA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
XX
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 310..6315
FT /*tag= a
XX
PN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
(NOVS) NOVARTIS AG.
PA ACN44182 Human gen
PA ACN44182 Mouse gen
XX
PI Adb71938 Rice gene
PI Adb31378 testoster
XX
XX Aak73030 Human imm
DR Aak73028 Human imm
DR Aax34746 DNA encod
XX
XX Aac68703 Ehrlichia
PT Abs63279 DNA encod
PS Aaa09740 E. chaffe
XX
XX Aav07179 Ehrlichia
XX Aac68716 Ehrlichia
XX Aas07578 DNA encod
CC ABL13734 Drosophil
CC ABL20392 Drosophil
CC ABL29345 Drosophil
CC ABL29344 Drosophil

Claim 6; Page 22-32; 48pp; English.
The present sequence is that of Arabidopsis thaliana ecotype Zurich
epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
protein (see AAB20062) with similarity with ARPase/helicase proteins of
the SWI2/SNF2 family, which controls gene silencing, particularly plant

CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression

SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 4; Length 6571;
Best Local Similarity 100.0%; Pred. No. 1.6e-147;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACAACATGAGTTTTCCTTCGCGTAATCGTAATCAAAATCGCTCGCGAAGAAC 60
DB 1 CACAACATGAGTTTTCCTTCGCGTAATCGTAATCAAAATCGCTCGCGAAGAAC 60
QY 61 GCGGTGCTAGAGAGAGAGTACCGGTACATATCTCGGAACCCAAATTCGTCTTCTT 120
DB 61 GCGGTGCTAGAGAGAGTACCGGTACATATCTCGGAACCCAAATTCGTCTTCTT 120
QY 121 CAAACTCAGTTTCCGAAACCCCAACACCGCGAGGATTGCGCTGAGAACACATT 180
DB 121 CAAACTCAGTTTCCGAAACCCCAACACCGCGAGGATTGCGCTGAGAACACATT 180
QY 181 AATCGAGAATTGCTGGGAATTCCTAAATTTTCCCTCGGTTTTTCTTTCACACTCTCG 240
DB 181 AATCGAGAATTGCTGGGAATTCCTAAATTTTCCCTCGGTTTTTCTTTCACACTCTCG 240
QY 241 AATCGGAATTTCCACCAAGCTCCGTCAAGCGATAGATTCTGCAATTAACACTTTTCG 300
DB 241 AATCGGAATTTCCACCAAGCTCCGTCAAGCGATAGATTCTGCAATTAACACTTTTCG 300
QY 301 GAGTGAATATGAAGAAAGATGAAGATGTTTTCACGGGAGAACCAATTTACACACGA 360
DB 301 GAGTGAATATGAAGAAAGATGAAGATGTTTTCACGGGAGAACCAATTTACACACGA 360
QY 361 TCCTAGCAGCTTCAATTTCTGCTCAGTTTGAACAAGAACCCCTGTTTGAGAGGTCA 420
DB 361 TCCTAGCAGCTTCAATTTCTGCTCAGTTTGAACAAGAACCCCTGTTTGAGAGGTCA 420
QY 421 AGCGGGGGACACCATCTAGAAAGGTAATACTCCAGCTTCTGCTACTAGAAAGTCAGAG 480
DB 421 AGCGGGGGACACCATCTAGAAAGGTAATACTCCAGCTTCTGCTACTAGAAAGTCAGAG 480
QY 481 AGACTGGCTCCCTCAGCTGC 500
DB 481 AGACTGGCTCCCTCAGCTGC 500

RESULT 2
AAA89385
ID AAA89385 standard; DNA; 10329 BP.
XX
AC AAA89385;
XX
DT 04-NOV-2004 (revised)
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Arabidopsis thaliana.
XX

FH	Key	Location/Qualifiers
FT	exon	1..1008
FT		/*tag= a
FT		/number= 1
FT	intron	1009..1295
FT		/*tag= b
FT		/number= 1
FT	exon	1296..2550
FT		/*tag= c
FT		/number= 2
FT	CDS	1301..9808
FT		/*tag= d
FT		/product= "Columbia protein"
FT		/note= "This sequence contains 15 introns."
FT	intron	2551..2673
FT		/*tag= e
FT		/number= 2
FT	exon	2674..2752
FT		/*tag= f
FT		/number= 3
FT	intron	2753..2867
FT		/*tag= g
FT		/number= 3
FT	exon	2868..3113
FT		/*tag= h
FT		/number= 4
FT	intron	3114..3506
FT		/*tag= i
FT		/number= 4
FT	exon	3507..3680
FT		/*tag= j
FT		/number= 5
FT	intron	3681..3973
FT		/*tag= k
FT		/number= 5
FT	exon	3974..4895
FT		/*tag= l
FT		/number= 6
FT	allele	4338
FT		/*tag= m
FT		/note= "ecotype Zurich"
FT	intron	4896..4975
FT		/*tag= n
FT		/number= 6
FT	exon	4976..5217
FT		/*tag= o
FT		/number= 7
FT	intron	5218..5777
FT		/*tag= p
FT		/number= 7
FT	exon	5778..5882
FT		/*tag= q
FT		/number= 8
FT	intron	5883..6082
FT		/*tag= r
FT		/number= 8
FT	exon	6083..7480
FT		/*tag= s
FT		/number= 9
FT	allele	6721
FT		/*tag= t
FT		/note= "ecotype Zurich"
FT	intron	7481..7615
FT		/*tag= u
FT		/number= 9
FT	exon	7616..7771
FT		/*tag= v
FT		/number= 10
FT	intron	7772..7914
FT		/*tag= w
FT		/number= 10
FT	exon	7915..8070
FT		/*tag= x

FT intron /number= 11
 FT 8071. .8153
 FT /*tag= y
 FT /number= 11
 FT 8154. .8318
 FT /*tag= z
 FT /number= 12
 FT 8319. .8451
 FT /*tag= aa
 FT /number= 12
 FT 8452. .8629
 FT /*tag= ab
 FT /number= 13
 FT 8630. .8718
 FT /*tag= ac
 FT /number= 13
 FT 8719. .8918
 FT /*tag= ad
 FT /number= 14
 FT 8919. .9000
 FT /*tag= ae
 FT /number= 14
 FT 9001. .9211
 FT /*tag= af
 FT /number= 15
 FT 9212. .9284
 FT /*tag= ag
 FT /number= 15
 FT 9285. .10329
 FT /*tag= ah
 FT /number= 16

XX WO20010801-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.

XX 23-JUN-1999; 99GB-00014623.

XX (NOVS) NOVARTIS AG.

XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;

XX WPI; 2001-137952/14.

XX P-PSDB; AAB20063.

XX Novel gene encoding a protein that controls gene silencing, in particular silencing of plant genes.

XX Claim 5; Page; 48pp; English.

XX The present sequence is that of an Arabidopsis thaliana ecotype Columbia
 CC gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino
 CC acid protein with similarity with ATPase/helicase proteins of the
 CC SWI2/SNF2 family, and which controls gene silencing, particularly plant
 CC gene silencing. Trans-acting modified loci were identified by T-DNA
 CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
 CC carries a heritable inactivated, methylated hygromycin resistance gene
 CC (hyg). Mutation of a silencing modified locus results in release of
 CC silencing of the hyg gene and restores hygromycin resistance. Plants
 CC homozygous for the silent resistance gene were subjected to
 CC transformation with a bar selectable marker gene under control of the T-
 CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
 CC screened for hygromycin resistance. The mutant phenotype (hygromycin
 CC resistance) was screened for genetic co-segregation with a specific T-DNA
 CC insert. Cloning of the tagged gene allowed characterization of the
 CC present sequence. 2 Allelic differences were found in the corresponding
 CC gene of A. thaliana ecotype Zurich, occurring at nucleotide positions
 CC 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful
 CC as a molecular tool for regulating gene expression. Note: the present
 CC sequence is not shown in the specification but is derived from the A.

CC thaliana mutant silencer gene sequence given in AAA89353
 CC
 CC Revised record issued on 04-NOV-2004 : Correction to feature table key
 XX
 SQ Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;
 Query Match 60.8%; Score 304; DB 4; Length 10329;
 Best Local Similarity 100.0%; Pred. No. 3.le-85;
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 CACAAGCATGAGTTTTTTCCTTCGGTAATCGTAAATCAAAAATCGCTCGCTCGAGAAAC 60
 Db 705 CACAAGCATGAGTTTTTTCCTTCGGTAATCGTAAATCAAAAATCGCTCGCTCGAGAAAC 764
 Qy 61 GCCGGTCTAGAAGAGGAAAGTACCGTACATAATCTCTGCGAACCCAAATTTCTCGTCTTT 120
 Db 765 GCCGGTCTAGAAGAGGAAAGTACCGTACATAATCTCTGCGAACCCAAATTTCTCGTCTTT 824
 Qy 121 CAAACTCAGTTTTTCCGAAACCCCAACACCGCGAGGATTGCGATGGCTGAAGAACCATT 180
 Db 825 CAAACTCAGTTTTTCCGAAACCCCAACACCGCGAGGATTGCGATGGCTGAAGAACCATT 884
 Qy 181 AATCGAGAAATGTGCTGGAAATTTCTCAAATTTCCCTCGCGTTTTTTTTCACACTCTCGG 240
 Db 885 AATCGAGAAATGTGCTGGAAATTTCTCAAATTTCCCTCGCGTTTTTTTTCACACTCTCGG 944
 Qy 241 AATCGGAAATTTCCACCAAGCTCCGTCAGCGATAGATTCTGCAATTTACACACTTTTCGC 300
 Db 945 AATCGGAAATTTCCACCAAGCTCCGTCAGCGATAGATTCTGCAATTTACACACTTTTCGC 1004
 Qy 301 GCAG 304
 Db 1005 GCAG 1008

RESULT 3

AAA89353

ID AAA89353 standard; DNA; 10329 BP.

XX AC AAA89353;

XX DT 04-NOV-2004 (revised)

XX DT 23-APR-2001 (first entry)

XX DE Arabidopsis thaliana gene involved in epigenetic gene silencing.

XX KW Gene silencing; silencing gene; MOM; ds.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT exon 1. .1008
 FT /*tag= a
 FT /number= 1
 FT intron 1009. .1295
 FT /*tag= b
 FT /number= 1
 FT exon 1296. .2550
 FT /*tag= c
 FT /number= 2
 FT CDS 1301. .9808
 FT /*tag= d
 FT /product= "MOM protein"
 FT /note= "This sequence contains 15 introns."
 FT intron 2551. .2673
 FT /*tag= e
 FT /number= 2
 FT exon 2674. .2752
 FT /*tag= f
 FT /number= 3
 FT intron 2753. .2867
 FT /*tag= g
 FT /number= 3

Qy	241	AATCGGAATTTCCACCAAGCTCCGTC	CAAGCATAGATTCTGCA	CAATTACACACTTTGCG	300
Db	945	AATCGGAATTTCCACCAAGCTCCGTC	CAAGCATAGATTCTGCA	CAATTACACACTTTGCG	1004
Qy	301	GCAG	304		
Db	1005	GCAG	1008		

RESULT 4
ACN44648
ID ACN44648 standard; DNA; 30277 BP.

DT 18-NOV-2004 (first entry)

DE Mouse genomic sequence mCG22518.

XX
KW
Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

OS Mus musculus.

XX PN WO2003073826-A2.

12-SEP-2003.

28-FEB-2003: 2003WO-US006235.

PR 01-MAR-2002: 2002US-00087192.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW:

XX
DR WPI: 2003-328604/31.

Recombinant nucleic acid useful for diagnosis and treatment of carcinoma comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1201; Opp; English.

The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US2002192586A1, for which no sequence data was published

SQ Sequence 30277 BP; 7585 A; 5986 C; 6063 G; 6356 T; 0 U; 4287 Other;

Query Match	7.3†	Score 36.4;	DB 11;	Length 30277;
Best Local Similarity	45.2‡	Pred. No. 4.1;		
Matches 133;	Conservative	0;	Mismatches 161;	Indels 0;
				Gaps 0;

QY 204 TCAAAATTTCCCTCGCGTTTCTTTCACTCTCGGAATCGGAAATTTCCACCAAGCTC 263

Db 17672 TTAACATATCTTCTGGCCTCCTTATGCACACTATGGCATGCAAAATGTATGCACCTTAGCA 17731

264 CGTCAAGCGATAGATTCTGACAAATTACACACTTTCGCCGAGTGAATATGAAGAAAGATGA 323

Db	17792	C	A	A	G	G	G	T	G	A	C	A	G	T	G	G	T	T	G	G	C	A	T	G	T	C	A	T	G	T	G	C	C	C	A	T	T	T	A	C	17851												
Qy	384	C	T	C	A	G	T	T	G	A	A	A	C	C	C	T	G	T	T	T	G	A	G	G	T	C	A	A	G	C	C	C	C	C	C	C	A	C	A	T	C	A	G	N	A	443							
Db	17852	C	T	C	A	T	A	G	C	C	A	T	C	A	A	C	G	T	G	C	A	T	G	C	A	T	G	C	A	A	A	T	A	T	A	C	A	G	A	T	A	C	A	G	C	T	G	17911					
Qy	444	G	G	T	A	A	T	A	A	C	T	T	C	T	G	T	C	T	A	G	A	A	A	G	T	C	A	G	A	G	A	C	T	G	G	T	C	C	C	T	C	A	C	497									
Db	17912	C	T	T	G	A	C	A	T	C	A	G	G	T	T	T	G	A	G	A	T	C	T	T	C	A	G	G	T	T	T	G	G	T	T	G	G	T	C	T	G	C	T	G	C	C	C	C	T	C	A	C	17965

RESULT 5
ACN44182/c
ID ACN44182 standard; DNA; 134738 BP.

AC ACN44182;

DT 18-NOV-2004 (first entry)

DE Human genomic sequence hCG37331.

XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.

OS Homo sapiens.

PN WO2003073826-A2.

PD 12-SEP-2003.

PF 28-FEB-2003; 2003WO-US006235.

PR 01-MAR-2002; 2002US-00087192.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW;

DR WPI; 2003-328604/31.

Recombinant nucleic acid useful for diagnosis and treatment of carcinoma PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 502; Opp; English.

The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CA coding sequence. Note: This patent is an equivalent to basic patent US20021182596A1, for which no sequence data was published.

SQ Sequence 134738 BP; 35840 A; 20570 C; 21056 G; 38635 T; 0 U; 18637 Other;
Query Match 7.3%; Score 36.4; DB 11; Length 134738;

Ov 276 GATTCTGCAAAATTACACATTTCCGCGCAGTGAATATGAAAGAAAGATGAAAGATTGGTTT 335
 Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0
 best local similarity 30.8%; Fied. NO. 7.6;

128041 GCTTCTGACTATGAGGTACAGGGCCATGGTGAATGCTTATATAACTTATTAGGGGATGT 127

336 GACGGGGAGAACCATTTACACCAGATCCCTAGCAGCTTCAATTCCCTGCCTCAGTTGAACA 395

Db 461 WTTWKSCAWCYTKSYTWRRAWSKCCCMRGSGCYKYMARS CCYKGRAMYTKRAAWT 402

Qy 289 ACA 291

Db :
401 WAA 399

RESULT 8

AAK73030/c

ID AAK73030 standard; DNA; 2339 BP.

XX

AC AAK73030;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27842.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001354.

XX

31-JAN-2000; 2000US-0179065P.

PR

04-FEB-2000; 2000US-0180628P.

PR

24-FEB-2000; 2000US-0184664P.

PR

02-MAR-2000; 2000US-0186350P.

PR

16-MAR-2000; 2000US-0189874P.

PR

17-MAR-2000; 2000US-0190076P.

PR

18-APR-2000; 2000US-0198123P.

PR

19-MAY-2000; 2000US-0205515P.

PR

07-JUN-2000; 2000US-0209467P.

PR

28-JUN-2000; 2000US-0214886P.

PR

30-JUN-2000; 2000US-0215135P.

PR

07-JUL-2000; 2000US-0216647P.

PR

07-JUL-2000; 2000US-0216880P.

PR

11-JUL-2000; 2000US-0217487P.

PR

11-JUL-2000; 2000US-0217496P.

PR

14-JUL-2000; 2000US-0218290P.

PR

26-JUL-2000; 2000US-0220963P.

PR

26-JUL-2000; 2000US-0220964P.

PR

14-AUG-2000; 2000US-0224518P.

PR

14-AUG-2000; 2000US-0224519P.

PR

14-AUG-2000; 2000US-0225213P.

PR

14-AUG-2000; 2000US-0225214P.

PR

14-AUG-2000; 2000US-0225266P.

PR

14-AUG-2000; 2000US-0225267P.

PR

14-AUG-2000; 2000US-0225268P.

PR

14-AUG-2000; 2000US-0225270P.

PR

14-AUG-2000; 2000US-0225447P.

PR

14-AUG-2000; 2000US-0225757P.

PR

14-AUG-2000; 2000US-0225758P.

PR

14-AUG-2000; 2000US-0225759P.

PR

18-AUG-2000; 2000US-0226279P.

PR

22-AUG-2000; 2000US-0226681P.

PR

22-AUG-2000; 2000US-0226868P.

PR

22-AUG-2000; 2000US-0227182P.

PR

23-AUG-2000; 2000US-0227009P.

PR

30-AUG-2000; 2000US-0228924P.

PR

01-SEP-2000; 2000US-0229287P.

PR

01-SEP-2000; 2000US-0229343P.

PR

01-SEP-2000; 2000US-0229344P.

PR

01-SEP-2000; 2000US-0229345P.

PR

05-SEP-2000; 2000US-0229509P.

PR

05-SEP-2000; 2000US-0229513P.

PR

06-SEP-2000; 2000US-0230437P.

PR

06-SEP-2000; 2000US-0230438P.

PR

08-SEP-2000; 2000US-0231242P.

PR

PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog; ss.
XX Ehrlichia chaffeensis.
OS
XX WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US019600.
XX PR 19-SEP-1997; 97US-0059353P.
XX PA (OHIS) UNIV OHIO STATE.
XX PI Rikihisa Y, Ohashi N;
XX WPI; 1999-354290/21.
DR P-PSDB; AAY06946.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and Ehrlichia
PT canis.
PS Disclosure; Fig 6A; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown in
CC AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs. (Updated
CC on 27-AUG-2003 to correct OS field.)
XX Sequence 861 BP; 281 A; 140 C; 169 G; 271 T; 0 U; 0 Other;
SQ

Query Match 6.8%; Score 33.8; DB 2; Length 861;
Best Local Similarity 49.2%; Pred. No. 6.2;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTCACGGGGAGAACCTTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGATATCTAAATCAATCAATACCTGC 583
QY 378 TCCTGCTCAGTTGAAACAAGAAACCCCTGGTTTGAGGAGGTCAAGCCGGGGACACCATC 437
Db 582 ACATATATAAGGAGAAAGGTATGCTTCACATTATTACATCATATAACATGCGTTTCAGCAT 523
QY 438 TAGGAAGGTAAATTAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
Db 522 AAATGATTTTATCAAGTAGTCTTCATTATTATTAGAAAGACAGACGCACTGCTGCACCATC 463
QY 498 T 498
Db 462 T 462

RESULT 11
AAC68703/C
ID AAC68703 standard; DNA; 861 BP.
XX AAC68703;
XX 02-MAR-2001 (first entry)
XX Ehrlichia chaffeensis VSA2 gene partial coding sequence.
XX Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdorf3; ds.
XX Ehrlichia chaffeensis.
OS

PN WO20005063-A2.
XX 02-NOV-2000.
XX 21-APR-2000; 2000WO-US010886.
XX PF 22-APR-1999; 99US-0130725P.
XX PR (UYFL) UNIV FLORIDA.
XX PA Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, Mcguire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX WPI; 2000-679675/66.
DR P-PSDB; AAB36186.
XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium.
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens.
XX Claim 4; Page 39-40; 63pp; English.
XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) gene of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
CC 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids
XX Sequence 861 BP; 281 A; 140 C; 168 G; 272 T; 0 U; 0 Other;
SQ

Query Match 6.8%; Score 33.8; DB 3; Length 861;
Best Local Similarity 49.2%; Pred. No. 6.2;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTCACGGGGAGAACCTTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGATATCTAAATCAATCAATACCTGC 585
QY 378 TCCTGCTCAGTTGAAACAAGAAACCCCTGGTTTGAGGAGGTCAAGCCGGGGACACCATC 437
Db 584 ACATATATAAGGAGAAAGGTATGCTTCACATTATTACATCATATAACATGCGTTTCAGCAT 525
QY 438 TAGGAAGGTAAATTAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
Db 524 AAATGATTTTATCAAGTAGTCTTCATTATTATTAGAAAGACAGACGCACTGCTGCACCATC 465
QY 498 T 498
Db 464 T 464

RESULT 12
ABS63279/C
ID ABS63279 standard; DNA; 861 BP.
XX ABS63279;
XX 15-NOV-2002 (first entry)
XX DNA encoding Ehrlichia chaffeensis outer membrane protein (OMP) #4.
KW Outer membrane protein; OMP; gene; ds; P30F; ehrlichiosis; infection.
XX Ehrlichia chaffeensis.
XX US2002120115-A1.
PN

XX	19-MAY-1999;	99US-00314701.	
PF			
XX	18-SEP-1998;	98US-0100843P.	
XX			
XX	(OHIS.) UNIV OHIO STATE RES FOUND.		
PA			
XX	Rikihisa Y, Ohashi N;		
XX			
PI	WPI; 2003-553952/52.		
XX	P-PSDB; ADA09741.		
DR			
XX			
XX	New isolated polynucleotide encoding outer membrane protein P30 of		
PT	Ehrlichia canis or its variant or fragment, useful for producing		
PT	Ehrlichia canis or Ehrlichia chaffeensis outer membrane proteins for		
PT	diagnosing and treating ehrlichiosis.		
XX			
XX	Disclosure; Fig 6; 105pp; English.		
PS			
XX			
XX	The invention relates to an isolated polynucleotide encoding a variant of		
CC	the outer membrane protein (OMP) P30 of Ehrlichia canis, an outer		
CC	membrane protein of E. canis , or an antigenic fragment of the E. canis		
CC	P30 protein, or comprising a sequence which is the complement of		
CC	nucleotide 99 through nucleotide 672 of ADA09764. Also disclosed are E.		
CC	canis and E. chaffeensis outer membrane proteins and their encoding		
CC	nucleic acids. The polynucleotides are useful for producing E. canis or		
CC	E. chaffeensis outer membrane protein, for designing hybridisation probes		
CC	for isolating and identifying cDNA and genomic clones encoding the OMP		
CC	or its allelic forms, for designing primers for PCR. The polypeptides		
CC	encoded by the polynucleotide is useful for diagnosing human ehrlichiosis		
CC	(Rocky mountain spotted fever) or canine ehrlichiosis. The present		
CC	sequence encodes an E. chaffeensis outer membrane protein.		
XX			
XX	Sequence 861 BP; 281 A; 140 C; 169 G; 271 T; 0 U; 0 Other;		
QY			
Query Match	6.8%;	Score 33.8;	DB 9; Length 861;
Best Local Similarity	49.2%;	Pred. No. 6.2;	
Matches	89;	Conservative	0; Mismatches 92; Indels 0; Gaps 0;
QY	318	AGATGAAGACATTTGGTTTGACGGGGAGAACCATTTACACCAGATCCCTAGCAGCTTCAAT	377
Db	642	ATAGAAATTTTAGGATTTATAGCTTCMAACATGGATCTAATATCAATACCAATACCTGC	583
QY	378	TCCTGCCTCAGTTTGAACAAGAAACCCCTGGTTTGAGGAGGTCAAGCCGGGGACACCATC	437
Db	582	ACATATATAGGAGAAAAGTGATGCCCTTCCACTTATFACATCATACATGCGTTTCAGCAT	523
QY	438	TACGAGGTTAATTAATCCAGCTTCTGCTACTAGAAAGTCAAGAGACTGGCTCCCTCACC	497
Db	522	AAATGATTTTATCAAGTAGTCCCTTCATTTATTAGAAAGACAGACGCACTGCCTGCACCATC	463
QY	498	T	498
Db	462	T	462
RESULT 14			
AAV07179/c			
ID	AAV07179	standard; DNA; 4683 BP.	
XX			
XX	AAV07179;		
XX			
DT	14-SEP-1998	(first entry)	
XX			
DE	Ehrlichia chaffeensis	VSA genomic locus.	
XX			
XX	MAP1 homologue; variable surface antigen; VSA1; VSA2; VSA3; VSA4; VSA5;		
KW	rickettsia; DNA vaccine; ss.		
XX			
OS	Ehrlichia chaffeensis.		
Key		Location/Qualifiers	
FT	CDS	3..134	

XX	29-AUG-2002.
PD	28-JAN-2002; 2002US-00059964.
XX	19-MAY-1999; 99US-00314701.
XX	(RIKI// RIKIHISA Y.
PA	(OHAS// OHASHI N.
XX	Rikihisa Y, Ohashi N;
PI	WPI; 2002-618954/66.
XX	P-PSDB; ABG77938.
DR	
XX	
PT	Isolated polynucleotide encoding an outer membrane protein of E.canis or
PT	E.chaffeensis used in the diagnosis of infection.
XX	
PS	Disclosure; Fig 6A; 49pp; English.
XX	
CC	The invention relates to an isolated polynucleotide encoding an outer
CC	membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC	in the diagnosis of infection. An infection such as human ehrlichiosis or
CC	canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC	patient, providing a polypeptide or mixture of polypeptides, contacting
CC	the sample with the polypeptide and assaying for the formation of a
CC	complex between antibodies in the serum sample and the polypeptide, where
CC	formation of a complex is indicative of infection with E. chaffeensis.
CC	This sequence represents DNA encoding an Ehrlichia outer membrane protein
CC	of the invention
XX	
SQ	Sequence 861 BP; 281 A; 140 C; 169 G; 271 T; 0 U; 0 Other;
	Query Match 6.8%; Score 33.8; DB 6; Length 861;
	Best Local Similarity 49.2%; Pred. No. 6.2;
	Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0
Qy	318 AGATGAAGAAGATTGGTTTGACGGGGAGAACCACTTTACACCAGATCCCTTAGCAGCTTCAAT 377
Dd	
	642 ATAAAGAAATTTTACGANTTTATAGCTTCAAACATGATGACTAAATCAATACCAATACCTGC 583
Qy	378 TCCTGCCCTCAGTTGAACAAGAAAACCCCTGGTTGTAGGAGGTCAGCCGGGGGACACCATC 437
Dd	
	582 ACATATATAAGGAGAAAAAGTATGCCCTTCACTTATTACATCAATAACATGGCTTCAGCAT 523
Qy	438 TACGAAGGTAAATAACTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCAC 497
Dd	
	522 AAATGATTTATCAAGTAGTCTCTTATTTATTAGAAAGACAGCGCACTGCTCGCACCATC 463
Qy	498 T 498
Dd	462 T 462
RESULT 13	
ADA09740/c	
ID	ADA09740 standard; DNA; 861 BP.
XX	
AC	ADA09740;
XX	
DT	06-NOV-2003 (first entry)
XX	
DE	E. chaffeensis DNA encoding outer membrane protein OMP-1D.
XX	
KW	ds; gene; outer membrane protein; circulating leukocyte;
KW	monocytic ehrlichiosis; Rocky Mountain spotted fever;
KW	canine ehrlichiosis.
OS	Ehrlichia chaffeensis.
XX	
PN	US6544517-B1.
XX	
PD	08-APR-2003.

```

FT FT /*tag= a
FT FT /note= "VSA1 partial gene"
FT FT 203. .212
FT FT /*tag= b
FT FT 226. .239
FT FT /*tag= c
FT FT 349. .354
FT FT /*tag= d
FT FT 375. .380
FT FT /*tag= e
FT FT 430. .434
FT FT /*tag= f
FT FT 439. .1299
FT FT /*tag= g
FT FT /note= "VSA2"
FT FT 1300. .1309
FT FT /*tag= h
FT FT 1324. .1333
FT FT /*tag= i
FT FT 1347. .1361
FT FT /*tag= j
FT FT /note= "G-rich region"
FT FT 1473. .1478
FT FT /*tag= k
FT FT 1499. .1504
FT FT /*tag= l
FT FT 1554. .1558
FT FT /*tag= m
FT FT 1563. .2399
FT FT /*tag= n
FT FT /note= "VSA3"
FT FT 2444. .2456
FT FT /*tag= o
FT FT 2465. .2477
FT FT /*tag= p
FT FT 2491. .2505
FT FT /*tag= q
FT FT /note= "G-rich region"
FT FT 2618. .2623
FT FT /*tag= r
FT FT 2644. .2649
FT FT /*tag= s
FT FT 2699. .2703
FT FT /*tag= t
FT FT 2708. .3550
FT FT /*tag= u
FT FT /note= "VSA4"
FT FT 3586. .3596
FT FT /*tag= v
FT FT 3586. .3596
FT FT /*tag= x
FT FT /note= "G-rich region"
FT FT 3610. .3622
FT FT /*tag= w
FT FT 3767. .3772
FT FT /*tag= y
FT FT 3788. .3793
FT FT /*tag= z
FT FT 3847. .3851
FT FT /*tag= aa
FT FT 3856. .4683
FT FT /*tag= ab
FT FT /note= "truncated VSA5 gene"
FT FT
XX PN WO9816554-A1.
XX PD 23-APR-1998.
XX XX
XX PF 17-OCT-1997; 97WO-US019044.
XX XX 17-OCT-1996; 96US-00733230.
XX PA (UYFL ) UNIV FLORIDA.

```

```

XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM;
XX WPI: 1998-251232/22.
DR P-PSDB; AAW51091, AAW51092, AAW51093, AAW51094, AAW51095, AAW51096,
DR AAW51097, AAW51098, AAW51099, AAW51100, AAW51101, AAW51102,
DR AAW51103, AAW51104, AAW51105.
XX Composition containing nucleic acid encoding rickettsial antigen - useful
PT for, e.g. stimulating protective immune response in humans or animals.
XX Claim 4; Fig 2A-B; 39pp; English.
XX This is the DNA sequence of a 4.6 kb genomic locus of Ehrlichia
CC chaffeensis that was obtained using a PCR cloning strategy based on
CC identifying genes homologous to the major antigenic protein MAP1 (see
CC AAW51088) of Cowdria ruminantium. It includes 5 very similar but non-
CC identical open reading frames (ORFs), of which ORF1 is a partial gene and
CC ORF5 is nearly complete but lacks 5-7 amino acid codons and a termination
CC codon (see AAW51091-95). Due to their similarity to MAP1 surface antigen
CC genes of C. ruminantium, the E. chaffeensis ORFs are designated variable
CC surface antigen (VSA) genes 1-5. A claimed composition comprises a
CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
CC that elicits a protective immune response against a rickettsial pathogen.
CC The nucleic acid is used, in human or veterinary medicine, in vaccines to
CC protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The
CC nucleic acids are also useful as probes to identify related sequences,
CC e.g. for identification of organisms and for diagnosing infection. Use of
CC nucleic acid vaccines avoids the problem of protein purification
CC associated with protein-based vaccines. The nucleic acid does not
CC replicate in the host but remains episomal and capable of expressing
CC polypeptide for at least 19 mth
XX SQ Sequence 4683 BP; 1590 A; 753 C; 818 G; 1522 T; 0 U; 0 Other;

Query Match 6.8%; Score 33.8; DB 2; Length 4683;
Best Local Similarity 49.2%; Pred. No. 13;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTGACGGGAGAACCACTTTACACGAGTCCTTAGCAGCTTCAAT 377
Db 1080 ATAGAAATTTGGGATTTATAGCTTCAAACATGATATAATCAATCAATACCTGCG 1021
QY 378 TCCTGCCTCAGTGAACAGAAACCCCTGGTTTGAGGAGGTCAAGCCGGGGACACCATC 437
Db 1020 ACATATATAAGGAGAAAGGTATGCTTCACTTATTACATCAATCAATCGCTTCAGCAT 961
QY 438 TAGGAGGTAATAACTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 960 AAATGATTTATCAAGTAGTCTTCTGTTTATTAGAAAGACAGACGACACTGCTGCACCATC 901
QY 498 T 498
Db 900 T 900

RESULT 15
AAC68716/c
ID AAC68716 standard; DNA; 4683 BP.
XX AC AAC68716;
XX 02-MAR-2001 (first entry)
XX Ehrlichia chaffeensis 28 kDa gene locus.
XX Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
XX vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; ihworf3;
XX 4hworf1; 18hworf1; 3gdorf3; ds.
XX Ehrlichia chaffeensis.
XX

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:04:38 ; Search time 114.348 Seconds
(without alignments)
7154.840 Million cell updates/sec

Title: US-10-018-929C-2_COPY_1_500

Perfect score: 500
Sequence: 1 cacaagcatgagttttctt.....agactggtccctcacctgc 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	34.2	6.8	1497	4	US-09-220-132-94
C 2	33.8	6.8	861	3	US-08-953-326-8
C 3	33.8	6.8	861	4	US-09-314-701-7
C 4	33.8	6.8	861	4	US-09-553-662-8
C 5	33.8	6.8	861	4	US-10-062-994-8
C 6	33.2	6.6	2801	3	US-08-747-221B-30
C 7	33.2	6.6	2801	3	US-08-747-221B-32
C 8	33.2	6.6	2801	3	US-09-005-051-30
C 9	33.2	6.6	2801	3	US-09-005-051-32
C 10	33.2	6.6	2801	4	US-09-403-942F-30
C 11	33.2	6.6	2801	4	US-09-403-942F-32
C 12	33.2	6.6	2836	3	US-08-747-221B-24
C 13	33.2	6.6	2836	3	US-08-747-221B-26
C 14	33.2	6.6	2836	3	US-09-005-051-24
C 15	33.2	6.6	2836	3	US-09-005-051-26
C 16	33.2	6.6	2836	4	US-09-403-942F-24
C 17	33.2	6.6	2836	4	US-09-403-942F-26
C 18	32.8	6.6	601	4	US-09-949-016-136606
C 19	32.8	6.6	162914	4	US-09-949-016-15578
C 20	32.8	6.6	246444	4	US-09-949-016-13113
C 21	32.6	6.5	3704	4	US-09-949-016-5607
C 22	32	6.4	31842	4	US-09-949-016-15123
C 23	31.6	6.3	552	4	US-09-270-767-3420
C 24	31.6	6.3	552	4	US-09-270-767-18702
C 25	31.2	6.2	193555	4	US-09-949-016-15553
C 26	31.2	6.2	193555	4	US-09-949-016-15554
C 27	31.2	6.2	193555	4	US-09-949-016-15555

ALIGNMENTS

RESULT 1

US-09-220-132-94/c

; Sequence 94, Application US/09220132

; Patent No. 6506607

; GENERAL INFORMATION:

; APPLICANT: Shivjan, Andrew W.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT

; FILE REFERENCE: 07334-074001

; CURRENT APPLICATION NUMBER: US/09/220,132

; CURRENT FILING DATE: 1998-12-23

; PRIOR APPLICATION NUMBER: US 60/079,303

; PRIOR FILING DATE: 1998-03-25

; PRIOR APPLICATION NUMBER: US 60/068,821

; PRIOR FILING DATE: 1997-12-24

; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 94

; LENGTH: 1497

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; LOCATION: (1)..(1497)

; OTHER INFORMATION: n = A,T,C or G

US-09-220-132-94

Query Match 6.8%; Score 34.2; DB 4; Length 1497;

Best Local Similarity 24.4%; Pred. No. 0.93; Mismatches 43; Indels 0; Gaps 0;

Matches 30; Conservative 50; Mismatches 43; Indels 0; Gaps 0;

QY 169 GAAGAACCATTATCGAATGTGCTGAATTCCTCAAAATTTCCCTCCGTTTCTT 228

Db 521 RAARRWTGKQWATWAAATTKRACMAAAKRWCMCMAMYYTTCMCAARRWTKRCMA 462

QY 229 TCACATCTCGGAATCGGAAATTTCCACCAAGCTCGTCAAGCGATAGATTTCTGCACAA 288

Db 461 WTTWKKSCAWCYTKSYTWRRAWKSCCCCMRGSCYKKNYARSCCYKGRAMYTKRAAAWT 402

QY 289 ACA 291

Db 401 WAA 399

RESULT 2

US-08-953-326-8/c

; Sequence 8, Application US/08953326

; Patent No. 6251872

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

```
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nvika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-8

Query Match      6.8%; Score 33.8; DB 3; Length 861;
Best Local Similarity 49.2%; Pred. No. 0.94;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGATACATAATCAATCAATACCTGCG 585

QY 378 TCCTGCTCAGTTTGAACAAGAAACCCCTGTTTGAGGAGGTCAAGCGGGGGACACCATC 437
Db 584 ACATATATAAGGAGAAAGGTATGCTTACATTTATACATCATACATGCGTTGAGCAT 525

QY 438 TAGGAAGGTAAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 524 AAATGATTTATCAAGTAGTCTTTCATTTATTAGAAAGACAGACGCACTGCTGCACCATC 465

QY 498 T 498
Db 464 T 464

RESULT 3
US-09-314-701-7/c
; Sequence 7, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(861)
US-09-314-701-7

Query Match      6.8%; Score 33.8; DB 4; Length 861;
Best Local Similarity 49.2%; Pred. No. 0.94;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
```

```
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAAAACATGGATCTAAATCAATACCAATACCTGCG 583
QY 378 TCCTGCTCAGTTTGAACAAGAAACCCCTGTTTGAGGAGGTCAAGCGGGGGACACCATC 437
Db 582 ACATATATAAGGAGAAAGGTATGCTTCACTTTATTACATCATACATGCGTTGAGCAT 523
QY 438 TAGGAAGGTAAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 522 AAATGATTTATCAAGTAGTCTTTCATTTATTAGAAAGACAGACGCACTGCTGCACCATC 463

QY 498 T 498
Db 462 T 462

RESULT 4
US-09-553-662-8/c
; Sequence 8, Application US/09553662
; Patent No. 6593147
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Allenman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; TITLE OF INVENTION: Methods of use
; FILE REFERENCE: UF-167XC3
; CURRENT APPLICATION NUMBER: US/09/553,662
; CURRENT FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-09-553-662-8

Query Match      6.8%; Score 33.8; DB 4; Length 861;
Best Local Similarity 49.2%; Pred. No. 0.94;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAGAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAAAACATGGATACATAATCAATACCAATACCTGCG 585
QY 378 TCCTGCTCAGTTTGAACAAGAAACCCCTGTTTGAGGAGGTCAAGCGGGGGACACCATC 437
Db 584 ACATATATAAGGAGAAAGGTATGCTTACATTTATACATCATACATGCGTTGAGCAT 525
QY 438 TAGGAAGGTAAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 524 AAATGATTTATCAAGTAGTCTTTCATTTATTAGAAAGACAGACGCACTGCTGCACCATC 465

QY 498 T 498
Db 464 T 464

RESULT 5
US-10-062-994-8/c
; Sequence 8, Application US/10062994
```

```

RESULT 6
US-08-747-221B-30
; Sequence 30, Application US/08747221B
; Patent No. 6063610
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: No. 6063610a1 Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:

```

```

1 GENERAL INFORMATION:
2
3 APPLICANT: Silver, Gary W.
4 APPLICANT: Wisniewski, Nancy
5
6 TITLE OF INVENTION: No. 6063610s1 Carboxylesterase Nucleic Acid
7
8 TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
9
10 NUMBER OF SEQUENCES: 66
11
12 CORRESPONDENCE ADDRESS:
13
14 ADDRESSEE: Carol Talkington Verser, Ph.D.
15 ADDRESSEE: Heska Corporation
16
17 STREET: 1825 Sharp Point Drive
18
19 CITY: Fort Collins
20
21 STATE: Colorado
22
23 COUNTRY: USA
24
25 ZIP: 80525
26
27 COMPUTER READABLE FORM:
28
29 MEDIUM TYPE: Floppy disk
30
31 COMPUTER: IBM PC compatible
32
33 OPERATING SYSTEM: Windows 95
34
35 SOFTWARE: Wordperfect for Windows, Version 7.0
36
37 CURRENT APPLICATION DATA:
38
39 APPLICATION NUMBER: US/08/747.221B
40

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2801 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..1886
;
US-09-005-051-30

Query Match 6.6%; Score 33.2; DB 3; Length 2801;
Best Local Similarity 52.7%; Pred. No. 2.8;
Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATCTCAAATTTTCCTCGCGGTTTTTCTTTCACACTCTCGGNATCGGAATTCGAAATTCACCA 258
2502 AATGTTGTAATTAACATACTATTTTTTATCTATCACACCGCTTAAATTGG----TGCAAAGA 2557

QY 259 AGCTCGTCAAGGATAGATTCTGACAAATTACACACTTTGCGGCAGTGAATATGAAGAAA 318
2558 AAAATTTTACTGTGATAATATTGACATTTACACAATATTACGAATTTGTAACCTCAAT 2617

QY 319 GATGAAAAGATTGGTTTGACGGGGGAGAACATTTTACCCAGATCCCTAGCAGCTTCAATT 378
2618 TAATGCAATATTGTTTTTTGTTTAAAAAACAATACATGACTTTTCTATATCAATTTTATATT 2677

QY 379 CCTG 382
2678 ACGG 2681

```

RESULT 9
US-09-005-051-32/c
; Sequence 32, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,051
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/747,221
; FILING DATE: No. 6291222ember 12, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: FC-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2801 nucleotides

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-09-005-051-32

Query Match 6.6%; Score 33.2; DB 3; Length 2801;

Best Local Similarity 52.7%; Pred. No. 2.8;

Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATTCTCAAAATTTCCCTCCGGTTTTCTTTTCACACTCTCGGAATCGAAATTTCCACCA 258

DB 300 AATGTTGTAATAATAACTATTTTATCTATCAACCGCTTAATTGG----TGCAGA 245

QY 259 AGCTCCGTCAGCGATAGATTCTGACAAATTACACACTTTCCGCGAGTGAATATGAAGAAA 318

DB 244 AAAATTTACTGTGATATATTTGACATTTTACACAATATTACGAATTTGTAACCTCAAT 185

QY 319 GATGAAAGATTTGGTTTGACGGGAGAACCAATTTACACAGATCCCTAGCAGCTTCAATT 378

DB 184 TATGTGAATATTGTTTTTGTGTAATAAAACATACATGACTTTTCTATATCATTTTATATT 125

QY 379 CCTG 382

DB 124 ACGG 121

RESULT 10

US-09-403-942F-30

; Sequence 30, Application US/09403942F

; Patent No. 6664090

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.

; APPLICANT: Brandt, Kevin S.

; APPLICANT: Wisniewski, Nancy

; TITLE OF INVENTION: NOVEL CARBOXYLESTERASE NUCLEIC ACID MOLECULES, PROTEINS AND USES

; FILE REFERENCE: FC-1-C1-PUS

; CURRENT APPLICATION NUMBER: US/09/403,942F

; CURRENT FILING DATE: 2000-05-02

; PRIOR APPLICATION NUMBER: PCT/US97/20598

; PRIOR FILING DATE: 1997-11-10

; PRIOR APPLICATION NUMBER: 08/747,221

; PRIOR FILING DATE: 1996-11-12

; NUMBER OF SEQ ID NOS: 76

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 30

; LENGTH: 2801

; TYPE: DNA

; ORGANISM: Ctenocephalides felis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (99)..(1886)

; OTHER INFORMATION:

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (2275)..(2275)

; OTHER INFORMATION: n = unknown

US-09-403-942F-30

Query Match 6.6%; Score 33.2; DB 4; Length 2801;

Best Local Similarity 52.7%; Pred. No. 2.8;

Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATTCTCAAAATTTCCCTCCGGTTTTCTTTTCACACTCTCGGAATCGAAATTTCCACCA 258

DB 2502 AATGTTGTAATAATAACTATTTTATCTATCAACCGCTTAATTGG----TGCAGA 2557

QY 259 AGCTCCGTCAGCGATAGATTCTGACAAATTACACACTTTCCGCGAGTGAATATGAAGAAA 318

DB 2558 AAAATTTACTGTGATATATTTGACATTTTACACAATATTACGAATTTGTAACCTCAAT 2617

QY 319 GATGAAAGATTTGGTTTGACGGGAGAACCAATTTACACAGATCCCTAGCAGCTTCAATT 378

DB 2618 TATGTGAATATTGTTTTTGTGTAATAAAACATACATGACTTTTCTATATCATTTTATATT 2677

QY 379 CCTG 382

DB 2678 ACGG 2681

RESULT 11

US-09-403-942F-32/c

; Sequence 32, Application US/09403942F

; Patent No. 6664090

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.

; APPLICANT: Brandt, Kevin S.

; APPLICANT: Wisniewski, Nancy

; TITLE OF INVENTION: NOVEL CARBOXYLESTERASE NUCLEIC ACID MOLECULES, PROTEINS AND USES

; FILE REFERENCE: FC-1-C1-PUS

; CURRENT APPLICATION NUMBER: US/09/403,942F

; CURRENT FILING DATE: 2000-05-02

; PRIOR APPLICATION NUMBER: PCT/US97/20598

; PRIOR FILING DATE: 1997-11-10

; PRIOR APPLICATION NUMBER: 08/747,221

; PRIOR FILING DATE: 1996-11-12

; NUMBER OF SEQ ID NOS: 76

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 32

; LENGTH: 2801

; TYPE: DNA

; ORGANISM: Ctenocephalides felis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (527)..(527)

; OTHER INFORMATION: n = unknown at position 527

US-09-403-942F-32

Query Match 6.6%; Score 33.2; DB 4; Length 2801;

Best Local Similarity 52.7%; Pred. No. 2.8;

Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATTCTCAAAATTTCCCTCCGGTTTTCTTTTCACACTCTCGGAATCGAAATTTCCACCA 258

DB 300 AATGTTGTAATAATAACTATTTTATCTATCAACCGCTTAATTGG----TGCAGA 245

QY 259 AGCTCCGTCAGCGATAGATTCTGACAAATTACACACTTTCCGCGAGTGAATATGAAGAAA 318

DB 244 AAAATTTACTGTGATATATTTGACATTTTACACAATATTACGAATTTGTAACCTCAAT 185

QY 319 GATGAAAGATTTGGTTTGACGGGAGAACCAATTTACACAGATCCCTAGCAGCTTCAATT 378

DB 184 TATGTGAATATTGTTTTTGTGTAATAAAACATACATGACTTTTCTATATCATTTTATATT 125

QY 379 CCTG 382

DB 124 ACGG 121

RESULT 12

US-08-747-221B-24

; Sequence 24, Application US/08747221B

; Patent No. 6063610

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.

; APPLICANT: Wisniewski, Nancy

; TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

; NUMBER OF SEQUENCES: 66

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carol Talkington Verser, Ph.D.

; ADDRESSEE: Heska Corporation

; STREET: 1825 Sharp Point Drive

```
/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/747,221B
/ FILING DATE: NO. 6063610ember 12, 1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ REFERENCE/DOCKET NUMBER: FC-1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 970/493-7272
/ TELEFAX: 970/484-9505
/ INFORMATION FOR SEQ ID NO: 24:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2836 nucleotides
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..1889
/ US-08-747-221B-24

Query Match 6.6%; Score 33.2; DB 3; Length 2836;
Best Local Similarity 52.7%; Pred. No. 2.8;
Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATTCTCAAATTTCCCTCGCGTTTTCTTTTCACACTCTCGGAATCGGAAATTTCCACCA 258
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2505 AATGTTGTAATTAACACTATTTTATCTATCAACACCGCTTAATTGG---TGCAGA 2560

QY 259 AGCTCGCTCAAGCGATAGATTCTGCACAAATACACATTTTCGCGCAGTGAATATGAAGAA 318
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2561 AAAATTTTACTGTGATAATATTGACATTTACACAATATTACGAATTTGTAACCTCAAT 2620

QY 319 GATGAAGAAGATTGGTTTGACGGGAGAACCAATTTACACAGATCCCTAGCAGCTTCAATT 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2621 TAATGCAATATTGTTTTTTGTTTAAACAAATACATGACTTTTCTATATCATTTATATT 2680

QY 379 CCTG 382
Db 2681 ACGG 2684

RESULT 13
US-08-747-221B-26/c
; Sequence 26, Application US/08747221B
; Patent No. 6063610
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
/ COMPUTER READABLE FORM: disk
/ MEDIUM TYPE: Floppy disk

/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/747,221B
/ FILING DATE: NO. 6063610ember 12, 1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ REFERENCE/DOCKET NUMBER: FC-1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 970/493-7272
/ TELEFAX: 970/484-9505
/ INFORMATION FOR SEQ ID NO: 24:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2836 nucleotides
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..1889
/ US-08-747-221B-26

Query Match 6.6%; Score 33.2; DB 3; Length 2836;
Best Local Similarity 52.7%; Pred. No. 2.8;
Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;

QY 199 AATTCTCAAATTTCCCTCGCGTTTTCTTTTCACACTCTCGGAATCGGAAATTTCCACCA 258
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2505 AATGTTGTAATTAACACTATTTTATCTATCAACACCGCTTAATTGG---TGCAGA 2560

QY 259 AGCTCGCTCAAGCGATAGATTCTGCACAAATACACATTTTCGCGCAGTGAATATGAAGAA 318
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2561 AAAATTTTACTGTGATAATATTGACATTTACACAATATTACGAATTTGTAACCTCAAT 2620

QY 319 GATGAAGAAGATTGGTTTGACGGGAGAACCAATTTACACAGATCCCTAGCAGCTTCAATT 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2621 TAATGCAATATTGTTTTTTGTTTAAACAAATACATGACTTTTCTATATCATTTATATT 2680

QY 379 CCTG 382
Db 2681 ACGG 2684

RESULT 14
US-09-005-051-24
; Sequence 24, Application US/09005051
; Patent No. 6291222
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary W.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
; TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/005,051
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/747,221
```

FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 2836 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
NAME/KEY: CDS
LOCATION: 99..1889
US-09-005-051-24

Query Match 6.6%; Score 33.2; DB 3; Length 2836;
Best Local Similarity 52.7%; Pred. No. 2.8;
Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;
QY 199 AATTCTCAAAATTTTCCTCGCGTTTTCTTTTCACACTCTCGGAATCGGAATTTCCACCA 258
Db 2505 AATGTTGTAAATTAATACTATTTTATCTATCAACACCGCTTAATTGG-----TGCAGA 2560
QY 259 AGCTCCGTCAAGCGATAGATTCTGACAATTACACACTTTTCGCGCAGTGAATATGAAGAA 318
Db 2561 AAAATTTTACTGTGATTAATTTTGACATTTTACACAATATTACGAATTTGAACTCAAT 2620
QY 319 GATGAAAGATTTGGTTTGACGGGGAGAACCAATTTACACCAGATCCCTAGCAGCTTCAAT 378
Db 2621 TATGTGAATATTGTTTTTGTGTTAAACAAACATACATGACTTTTCTATATCATTTTATTT 2680
QY 379 CCTG 382
Db 2681 ACGG 2684

RESULT 15

US-09-005-051-26/c
Sequence 26, Application US/09005051
Patent No. 6291222
GENERAL INFORMATION:
APPLICANT: Silver, Gary W.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: No. 6291222el Carboxylesterase Nucleic Acid
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carol Talkington Verser, Ph.D.
ADDRESSEE: Hesk Corporation
STREET: 1825 Sharp Point Drive
CITY: Fort Collins
STATE: Colorado
COUNTRY: USA
ZIP: 80525
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,051
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/747,221
FILING DATE: No. 6291222ember 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-1
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 2836 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-005-051-26

Query Match 6.6%; Score 33.2; DB 3; Length 2836;
Best Local Similarity 52.7%; Pred. No. 2.8;
Matches 97; Conservative 0; Mismatches 83; Indels 4; Gaps 1;
QY 199 AATTCTCAAAATTTTCCTCGCGTTTTCTTTTCACACTCTCGGAATCGGAATTTCCACCA 258
Db 332 AATGTTGTAAATTAATACTATTTTATCTATCAACACCGCTTAATTGG-----TGCAGA 277
QY 259 AGCTCCGTCAAGCGATAGATTCTGACAATTACACACTTTTCGCGCAGTGAATATGAAGAA 318
Db 276 AAAATTTTACTGTGATTAATTTTGACATTTTACACAATATTACGAATTTGAACTCAAT 217
QY 319 GATGAAAGATTTGGTTTGACGGGGAGAACCAATTTACACCAGATCCCTAGCAGCTTCAAT 378
Db 216 TATGTGAATATTGTTTTTGTGTTAAACAAACATACATGACTTTTCTATATCATTTTATTT 157
QY 379 CCTG 382
Db 156 ACGG 153

Search completed: July 12, 2005, 10:17:58
Job time : 122.348 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 542.443 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929c-2_COPY_1_500

Perfect score: 500
Sequence: 1 cacaagcatgagttttctc.....agactggctccctcaactgc 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	7.6	1288	20	US-10-739-930-1571
2	36.4	7.3	30277	13	Sequence 1571, Ap
3	36.4	7.3	134738	13	Sequence 1201, Ap
4	34.6	6.9	1190	19	Sequence 502, App
5	34.2	6.8	1497	21	Sequence 87233, A
6	33.8	6.8	861	13	Sequence 94, Appl
7	33.8	6.8	861	13	Sequence 8, Appl

c 8	33.8	6.8	861	13	US-10-062-994-8	Sequence 8, Appl
c 9	33.8	6.8	861	15	US-10-314-639-7	Sequence 7, Appl
c 10	33.8	6.8	861	19	US-10-722-077-8	Sequence 8, Appl
c 11	33.8	6.8	861	20	US-10-901-714-7	Sequence 7, Appl
c 12	33.8	6.8	861	20	US-10-901-774-7	Sequence 7, Appl
c 13	33.8	6.8	1897	19	US-10-437-963-22860	Sequence 22860, A
c 14	33.8	6.8	4557	19	US-10-437-963-10388	Sequence 10388, A
c 15	33.4	6.7	566	14	US-10-184-644-56	Sequence 56, Appl
c 16	33.4	6.7	566	14	US-10-184-634-56	Sequence 56, Appl
c 17	33.4	6.7	927	18	US-10-424-599-136122	Sequence 136122, A
c 18	33.2	6.6	2801	18	US-10-678-521-30	Sequence 30, Appl
c 19	33.2	6.6	2801	18	US-10-678-521-32	Sequence 32, Appl
c 20	33.2	6.6	2836	18	US-10-678-521-24	Sequence 24, Appl
c 21	33.2	6.6	2836	18	US-10-678-521-26	Sequence 26, Appl
c 22	33.2	6.6	54355	11	US-09-997-722-19	Sequence 19, Appl
c 23	33	6.6	806	20	US-10-363-345A-15863	Sequence 15863, A
c 24	33	6.6	806	20	US-10-363-345A-15864	Sequence 15864, A
c 25	33	6.6	806	21	US-10-363-483A-15863	Sequence 15863, A
c 26	33	6.6	806	21	US-10-363-483A-15864	Sequence 15864, A
c 27	32.8	6.6	3675	19	US-10-437-963-36018	Sequence 36018, A
c 28	32.8	6.6	175189	21	US-10-741-600-17738	Sequence 17738, A
c 29	32.8	6.6	255439	20	US-10-719-993-6799	Sequence 6799, Ap
c 30	32.6	6.5	390	13	US-10-027-632-71732	Sequence 71732, A
c 31	32.6	6.5	390	13	US-10-027-632-312521	Sequence 312521, A
c 32	32.6	6.5	390	17	US-10-027-632-71732	Sequence 71732, A
c 33	32.6	6.5	390	17	US-10-027-632-312521	Sequence 312521, A
c 34	32.6	6.5	478	10	US-09-918-995-12637	Sequence 12637, A
c 35	32.6	6.5	2022	18	US-10-606-320-63	Sequence 63, Appl
c 36	32.6	6.5	2022	18	US-10-606-320-65	Sequence 65, Appl
c 37	32.6	6.5	2022	20	US-10-746-914-63	Sequence 63, Appl
c 38	32.6	6.5	2022	20	US-10-746-914-65	Sequence 65, Appl
c 39	32.6	6.5	6793	16	US-10-148-351-43	Sequence 43, Appl
c 40	32.6	6.5	7153	16	US-10-148-351-44	Sequence 44, Appl
c 41	32.4	6.5	201	21	US-10-741-600-64019	Sequence 64019, A
c 42	32.4	6.5	1071	9	US-09-938-842A-1038	Sequence 1038, Ap
c 43	32.4	6.5	1071	11	US-09-938-842A-1038	Sequence 1038, Ap
c 44	32.4	6.5	44362	21	US-10-741-600-17870	Sequence 17870, A
c 45	32.2	6.4	2000	9	US-09-938-842A-4289	Sequence 4289, Ap

ALIGNMENTS

RESULT 1
US-10-739-930-1571
; Sequence 1571, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 1571
; LENGTH: 1288
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; OTHER INFORMATION: Clone ID: BRANA-23APR03-CLUSTER7272_1
US-10-739-930-1571

Query Match 7.6%; Score 38; DB 20; Length 1288;
Best Local Similarity 53.3%; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 70;
Qy 173 AACCACTTAATCAGAGAAATGCTGGAATTTCTCAAAATTTTCCTCGGTTTTTCTTCAC 232
Db 1130 AAGAAATTTGCTGGGTATAGGCTGCTGATAGTAACTTTGGTTCTGAAATTTTGT 1189
Qy 233 ACTCTCGGAATCGGAATTTTCCACCAAGCTCCCTCAAGCGATAGATCTGCACAAATAC 292

Db 1190 GGTCTCTTACCTTTCTGCTCTTACCAATCTTTGTAAATCCGATTCGATTAATGGCCC 1249
QY 293 ACTTTCGGCAGTGAATATGAAGAAGATG 322
Db 1250 GATTTCACCTAGTAAAAAATAAAG 1279

RESULT 2

US-10-087-192-1201
; Sequence 1201, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1201
; LENGTH: 30277
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(30277)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1201

Query Match 7.3%; Score 36.4; DB 13; Length 30277;

Best Local Similarity 45.2%; Pred. No. 2.6;
Matches 133; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 204 TCAAAATTTCCCTCGGCTTTTCTTTCACACTCTCGGAATCGGAATTTCCACCAAGCTC 263
Db 17672 TTAACATATCTTCTGGCTCCTTATGCACACTATGGCATGCAATGTATGCATTTAGCA 17731
QY 264 CGTCAAGCGATAGTCTTGCAATTTACACACTTTTCGGCGAGTGAATATGAAGAAAGATGA 323
Db 17732 CACACACACACATACATACATACACGAATTTAGGAAAAGAGTTTTTAAAAATATAA 17791
QY 324 AAGATTTGTTTACGGGAGAACCATTTACACAGATCCCTAGCAGCTTCAATTCCTGC 383
Db 17792 CAAGAGGGTGTGACCAAGTGGCATGGGTTTGGCCATGTGTCCACTGGCCCATTTTACC 17851
QY 384 CTCAGTTGAACAGAAACCCCTGTTTGGAGAGTCAAGCCGGGAGACCATCTACGAA 443
Db 17852 CTCTATAGCCATCAACCTGTCATGCAGAGCCTCAGGAAGATACAGATCGGGTG 17911
QY 444 GGTAATAACTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 17912 CTTGACATCTACGGCTTTGAGATCTTCCAGGTTTGGCTGTGGCTGCCCTCACC 17965

RESULT 3

US-10-087-192-502/c
; Sequence 502, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 502
; LENGTH: 134738
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(134738)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-502

Query Match 7.3%; Score 36.4; DB 13; Length 134738;

Best Local Similarity 50.6%; Pred. No. 5.9;
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 276 GATTTCACAATTACACACTTTTCGGCGAGTGAATATGAAGAAGATGAAGAAGATTGGTTT 335
Db 128041 GCTTCTGACTATGAGGTACAGGCCCATGCTGAATGCTTATATAACTTATTAGGGGATGT 127982
QY 336 GACGGGAGAACCAATTACACAGATCCCTAGCAGCTTCAATTCCTGCCTCAGTTGAACA 395
Db 127981 GTTGGGTGAAGTCTAGCCCTCGCTTCATTGCTGTTCTTGTGGAACGCAATTATCAGA 127922
QY 396 AGAAACCCCTGGTTTGGAGAGGTCAAGCCGGGGACACCATCTTACGAAGGTAAT 449
Db 127921 AGATATGCTCATTTGTAGGGCTACATCCATGTAACCTTCTACACATGAACT 127868

RESULT 4

US-10-437-963-87233
; Sequence 87233, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 87233
; LENGTH: 1190
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_86199C.1
US-10-437-963-87233

Query Match 6.9%; Score 34.6; DB 19; Length 1190;

Best Local Similarity 51.6%; Pred. No. 1.8;
Matches 79; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 345 AACCAATTACACCAGATCCCTAGCAGCTTCAATTCCTGCCTCAGTTGAACAAGAAACCCC 404
Db 229 AACCAATGCATCGTCATGTGCCCCACGCTGCAAGACCTTCTGCAATGTGCGACTTCTACCCC 288
QY 405 TGGTTTGAAGGTCAAGCCGGGGGACACCATCTACAGGTAATTAACCTCCAGCTTCTGC 464
Db 289 GCGTGTCTGTGGCGGACCCCGGTTTACCGGGCGGACGCGCAACAACTTCTACTTCCAC 348
QY 465 TACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497

Db 349 GGCAGAGGACGACGACTTCTGCATCGTCTCC 381

RESULT 5

US-10-831-704-94/c
; Sequence 94, Application US/10831704
; Publication No. US20050100931A1
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND
; TITLE OF INVENTION: ASSESSMENT
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/10/831,704
; CURRENT FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US/10/155,653
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US/09/220,132
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1497)
; OTHER INFORMATION: n = A,T,C or G
US-10-831-704-94

Query Match 6.8%; Score 34.2; DB 21; Length 1497;

Best Local Similarity 24.4%; Pred. No. 2.7; Mismatches 43; Indels 0; Gaps 0;
Matches 30; Conservative 50;

QY 169 GAAGAACCACTTAATCGAGAAATGTGCTGGAATTCCTCAAAATTTCCCTCGGTTTCTT 228

Db 521 RAARRWTGGKNAWTAAATWKAARCAARAAKWMCCMAYTTYCMMGARRWTKRCMA 462

QY 229 TCACACTCTCGAATCGGAATTTCCACCAAGTCTCGTCAAGCATAGATTCGACAAAT 288

Db 461 WTTWKSANWCYTKSYTWRRWKSCCCNRGSGCYKKYWARSCCYKGRAMWTKRAAAAT 402

QY 289 ACA 291

Db 401 WAA 399

RESULT 6

US-10-062-994-8/c
; Sequence 8, Application US/10062994
; Publication No. US20010001661A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmore, William M.
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; TITLE OF INVENTION: Methods of use
; FILE REFERENCE: UP-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662

; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-8

Query Match 6.8%; Score 33.8; DB 13; Length 861;

Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAAGATTGGTTTGACGGGGAGAACCAATTTACACAGATCCCTAGACGTTCAAT 377

Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGATCTAAATCAATACCAATACCTGC 585

QY 378 TCCTGCCTCAGTTGAACAGAAACCCCTGTTTGAGGAGGTCAAGCCGGGGACACCATC 437

Db 584 ACATATATATAGGAGAAAGGTATGCCTTCACCTATTATACATATAACATCGCTTACGAT 525

QY 438 TAGCAAGGTAAATAACTCCAGCTTCTGCTACTAGAAAGTCAAGAGAGCTGGCTCCCTCACC 497

Db 524 AATGATTTATCAAGTAGTCCCTTCCTTCATTTATTAGAGACAGACGCACTGCTGCACCATC 465

QY 498 T 498

Db 464 T 464

RESULT 7

US-10-059-964-7/c
; Sequence 7, Application US/10059964
; Publication No. US20020120115A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. US20020120115A1d
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/059,964
; CURRENT FILING DATE: 2002-01-28
; EARLIER APPLICATION NUMBER: 09/314,701
; EARLIER FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(861)
US-10-059-964-7

Query Match 6.8%; Score 33.8; DB 13; Length 861;

Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAAGATTGGTTTGACGGGGAGAACCAATTTACACAGATCCCTAGACGTTCAAT 377

Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGATCTAAATCAATACCAATACCTGC 583

QY 378 TCCTGCCTCAGTTGAACAGAAACCCCTGTTTGAGGAGGTCAAGCCGGGGACACCATC 437

Db 582 ACATATATATAGGAGAAAGGTATGCCTTCACCTATTATACATATAACATCGCTTACGAT 523

QY 438 TAGAAGGTAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 522 AAATGATTTATCAAGTAGCTCTTCATTTATTAGAAAGACAGAGCACTGGCTGCACCATC 463
QY 498 T 498
Db 462 T 462

RESULT 8

US-10-062-994-8/c
; Sequence 8, Application US/10062994
; Publication No. US20020132789A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE OF INVENTION: Methods of use
; FILE REFERENCE: UP-167XC3
; CURRENT APPLICATION NUMBER: US/10/062,994
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-062-994-8

Query Match 6.8%; Score 33.8; DB 13; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGTACTAAATCAATCAACCATACCTGC 585
QY 378 TCCTGCTCAGTTGAACAAGAACCCCTGGTTTGAGGAGGTCAAGCGGGGGACACCATC 437
Db 584 ACATATATAAGGAGAAAAGGTATGCTTCACTTATTACATATAACATGCGTTGACGAT 525
QY 438 TAGAAGGTAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 524 AAATGATTTATCAAGTAGCTCTTCATTTATTAGAAAGACAGAGCACTGGCTGCACCATC 465
QY 498 T 498
Db 464 T 464

RESULT 9

US-10-314-639-7/c
; Sequence 7, Application US/10314639
; Publication No. US2003010391A1
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. US2003010391A110

; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/10/314,639
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(861)
US-10-314-639-7

Query Match 6.8%; Score 33.8; DB 15; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGTACTAAATCAATCAATACCATACCTGC 583
QY 378 TCCTGCTCAGTTGAACAAGAACCCCTGGTTTGAGGAGGTCAAGCGGGGGACACCATC 437
Db 582 ACATATATAAGGAGAAAAGGTATGCTTCACTTATTACATATAACATGCGTTGACGAT 523
QY 438 TAGAAGGTAATACCTCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACC 497
Db 522 AAATGATTTATCAAGTAGCTCTTCATTTATTAGAAAGACAGAGCACTGGCTGCACCATC 463
QY 498 T 498
Db 462 T 462

RESULT 10

US-10-722-077-8/c
; Sequence 8, Application US/10722077
; Publication No. US20040126871A1
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Bowie, Michael V.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Mahan, Suman M.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Moreland, Annie L.
; APPLICANT: Simbi, Bigboy H.
; APPLICANT: Whitmire, William M.
; APPLICANT: Alleman, Arthur R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases and
; FILE OF INVENTION: Methods of use
; FILE REFERENCE: UP-167XC3
; CURRENT APPLICATION NUMBER: US/10/722,077
; CURRENT FILING DATE: 2003-11-24
; PRIOR APPLICATION NUMBER: US/10/062,994
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/533,662
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/337,827
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 08/953,326
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 08/733,230
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 861

```
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-722-077-8

Query Match      6.8%; Score 33.8; DB 19; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
    |||||
Db 644 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGGATCTAAATCAATCAATACCTGC 585

QY 378 TCCTGCCTCAGTTGACAGAAACCCCTGGTTTGAGAGAGTCAAGCGGGGGACACCATC 437
    |||||
Db 584 ACATATATAAGGAGAAAGGTATGCTTCACTTATTACATCAATAACATGCGTTTCAGCAT 525

QY 438 TAGCAGGTAATAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
    |||||
Db 524 AAATGATTATCAAGTAGTCTTTCATTTATTAGAAAGACAGAGCGACTGCTGCACCATC 465

QY 498 T 498
    |
Db 464 T 464

RESULT 11
US-10-901-714-7/c
; Sequence 7, Application US/10901714
; Publication No. US20040265334A1
; GENERAL INFORMATION:
; APPLICANT: RIKIHISA, YASUKO
; APPLICANT: OHASHI, NORIO
; TITLE OF INVENTION: OUTER MEMBRANE PROTEIN OF EHRlichia CANIS AND EHRlichia
; TITLE OF INVENTION: CHAFFEENSIS
; FILE REFERENCE: 22727-04109
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: 09/314,701
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: 60/100,843
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-901-714-7

Query Match      6.8%; Score 33.8; DB 20; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
    |||||
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGGATCTAAATCAATCAATACCTGC 583

QY 378 TCCTGCCTCAGTTGACAGAAACCCCTGGTTTGAGAGAGTCAAGCGGGGGACACCATC 437
    |||||
Db 582 ACATATATAAGGAGAAAGGTATGCTTCACTTATTACATCAATAACATGCGTTTCAGCAT 523

QY 438 TAGCAGGTAATAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
    |||||
Db 522 AAATGATTATCAAGTAGTCTTTCATTTATTAGAAAGACAGAGCGACTGCTGCACCATC 463

QY 498 T 498
    |
Db 462 T 462

RESULT 12
US-10-901-774-7/c
; Sequence 7, Application US/10901774
; Publication No. US20040265334A1
; GENERAL INFORMATION:
; APPLICANT: RIKIHISA, YASUKO
; APPLICANT: OHASHI, NORIO
; TITLE OF INVENTION: OUTER MEMBRANE PROTEIN OF EHRlichia CANIS AND EHRlichia
; TITLE OF INVENTION: CHAFFEENSIS
; FILE REFERENCE: 22727-04109
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: 09/314,701
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: 60/100,843
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-901-774-7

Query Match      6.8%; Score 33.8; DB 20; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
    |||||
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGGATCTAAATCAATCAATACCTGC 583

QY 378 TCCTGCCTCAGTTGACAGAAACCCCTGGTTTGAGAGAGTCAAGCGGGGGACACCATC 437
    |||||
Db 582 ACATATATAAGGAGAAAGGTATGCTTCACTTATTACATCAATAACATGCGTTTCAGCAT 523

QY 438 TAGCAGGTAATAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
    |||||
Db 522 AAATGATTATCAAGTAGTCTTTCATTTATTAGAAAGACAGAGCGACTGCTGCACCATC 463

QY 498 T 498
    |
Db 462 T 462

RESULT 13
US-10-437-963-22860/c
; Sequence 22860, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT FILING DATE: 2003-05-14
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 22860
; LENGTH: 1897
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_27997C.1
US-10-437-963-22860

Query Match      6.8%; Score 33.8; DB 19; Length 1897;
Best Local Similarity 49.7%; Pred. No. 4.3;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
```

```
; Publication No. US20040265334A1
; GENERAL INFORMATION:
; APPLICANT: RIKIHISA, YASUKO
; APPLICANT: OHASHI, NORIO
; TITLE OF INVENTION: OUTER MEMBRANE PROTEIN OF EHRlichia CANIS AND EHRlichia
; TITLE OF INVENTION: CHAFFEENSIS
; FILE REFERENCE: 22727-04109
; CURRENT APPLICATION NUMBER: US/10/901,774
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: 09/314,701
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: 60/100,843
; PRIOR FILING DATE: 1998-09-18
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 7
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Ehrlichia chaffeensis
US-10-901-774-7

Query Match      6.8%; Score 33.8; DB 20; Length 861;
Best Local Similarity 49.2%; Pred. No. 2.7;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 318 AGATGAAAGATTGGTTTGACGGGGAGAACCATTTACACGAGATCCCTAGCAGCTTCAAT 377
    |||||
Db 642 ATAAGAAATTTTAGGATTTATAGCTTCAACATGGGATCTAAATCAATCAATACCTGC 583

QY 378 TCCTGCCTCAGTTGACAGAAACCCCTGGTTTGAGAGAGTCAAGCGGGGGACACCATC 437
    |||||
Db 582 ACATATATAAGGAGAAAGGTATGCTTCACTTATTACATCAATAACATGCGTTTCAGCAT 523

QY 438 TAGCAGGTAATAATCCAGCTTCTGCTACTAGAAAGTCAGAGAGACTGCTCCCTCACC 497
    |||||
Db 522 AAATGATTATCAAGTAGTCTTTCATTTATTAGAAAGACAGAGCGACTGCTGCACCATC 463

QY 498 T 498
    |
Db 462 T 462
```

```

, TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
,
, TITLE OF INVENTION: ACIDS ENCODING THE SAME
,
, FILE REFERENCE: P3430RIC227
,
, CURRENT APPLICATION NUMBER: US/10/184,644
,
, CURRENT FILING DATE: 2002-06-28
,
, Prior Application removed - See File Wrapper or Palm
,
, NUMBER OF SEQ ID NOS: 612
,
, SEQ ID NO 56
,
, LENGTH: 566
,
, TYPE: PRT
,
, ORGANISM: Homo Sapien
,
, ORGANISM: Homo Sapien
,
, US-10-184-644-56
,

```

[illegible]

Search completed: July 12, 2005, 11:12:30
Job time : 545.443 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:03:44 ; Search time 2463.72 Seconds
(without alignments)
7724.974 Million cell updates/sec

Title: US-10-018-929c-2_COPY_1_500

Perfect score: 500
Sequence: 1 cacaagcatgagtttctc.....agactggctccctcaactgc 500

Scoring table: IDENTITY NUC

Gapex 10.0, Gapex 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

1: gb_est1.*

2: gb_est2.*

3: gb_hic.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gss1.*

9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.6	40.1	769	8	AQ956408 LERAJ80TR
2	165.6	33.1	261	7	Z18494 ATTS0870 Gi
3	108	21.6	934	8	B10852 F3E13-T7 IG
4	101	20.2	256	9	CNS00N7X
5	99.8	20.0	719	8	AL081579 Arabidops
6	88.4	17.7	709	8	AQ956407 LERAJ80TR
7	72.6	14.5	684	8	BH981035 OGG49F03
8	69.4	13.9	512	8	BZ465288 BOMRU74TR
9	59.4	11.9	926	8	BH707844 BOMHE37TF
10	38.2	7.6	273	7	B12496 F18C19-Sp6
11	38.2	7.6	541	4	CR464122 CR464122
12	37.2	7.4	531	8	BI503269 BI70008B
13	37	7.4	553	6	AQ807478 HS_3207 B
14	36.8	7.4	342	7	CB404139 OSTRO18E9
15	36.6	7.3	542	9	T38213 EST103627 S
16	36.6	7.3	579	9	CG842963 Ynhw3577
17	36.4	7.3	443	9	CR831390 GROAA864A
18	36.4	7.3	487	9	CR310963 Medicago
19	36.4	7.3	782	9	CR513493 Medicago
20	36.4	7.3	867	8	AZ540768 ENTDT3TR
21	36.4	7.3	1037	8	AZ536287 ENTDT09TF
22	36.2	7.2	3667	3	AK036840 Mus muscu
23	36	7.2	376	8	AZ292909 RPCI-23-1
24	35.4	7.1	512	8	AQ611314 HS_5088_B

C	25	35.2	7.0	525	1	AI992163	wt75d01.x
	26	35.2	7.0	881	5	BU601218	AGENCOURT
	27	35.2	7.0	902	8	AZ535986	ENTCJ64TR
	28	35.2	7.0	925	8	AZ538777	ENTFO18TF
	29	35.2	7.0	976	8	BH158282	ENTSS071TR
	30	35	7.0	632	9	AG578249	Mus muscu
	31	35	7.0	1034	2	BP208819	601872313
	32	34.8	7.0	1029	8	CC274196	CH261-132
	33	34.6	6.9	529	5	BP174756	BP174756
	34	34.6	6.9	611	2	BE433453	EST399982
	35	34.6	6.9	672	8	AQ317042	CITBI-E1-
	36	34.6	6.9	692	6	CA087382	SCOGAM211
	37	34.6	6.9	777	8	BZ606985	WHABU15TF
	38	34.6	6.9	993	9	CL960565	OSIFCC004
	39	34.4	6.9	405	5	BP664053	BP664053
	40	34.4	6.9	827	6	CB096655	AF53-RPf
	41	34.2	6.8	485	9	CL881647	abf50f12-
	42	34.2	6.8	903	8	BZ993728	PUDFC42TD
	43	34.2	6.8	1415	9	CG744998	P037-4-B0
	44	34	6.8	281	2	BH146723	BH146723
	45	34	6.8	361	1	AV815318	AV815318

ALIGNMENTS

RESULT 1
AQ956408 769 bp DNA linear GSS 28-JAN-2000
LOCUS LERAJ80TR LERA Arabidopsis thaliana genomic clone LERAJ80, genomic survey sequence.

ACCESSION AQ956408

VERSION AQ956408.1 GI:6784109

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.

REFERENCE 1 (bases 1 to 769)

AUTHORS Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T., Feldlyum, T., Liang, F., Creasy, T. and Fraser, C.M.

TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms

JOURNAL Unpublished (2000)

COMMENT Contact: Xiaoying Lin

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: at@tigr.org

For additional information, see <http://www.tigr.org/cdb/at/at.html>

Seq primer: TR

Class: shotgun.

Location/Qualifiers

source

1..769

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Landsberg erecta"

/db_xref="taxon:3702"

/clone_lib="LERA"

/notes="Organ: Leaf; Vector: phos1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."

ORIGIN

Query Match 40.1%; Score 200.6; DB 8; Length 769;

Best Local Similarity 98.1%; Pred. No. 1.4e-50;

Matches 203; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 294 CTTTCGGCGTGAATGAGAAAGATGTTTACCGGGAGACCATTTA 353

Db 96 CTGCTCTCAGTGAATATGAAGAAGATGAAAAGATTGGTTTGACCGGGAGAACCAATTGA 155
 QY 354 CACCAGATCCCTAGCAGCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 413
 Db 156 CACCAGATCCCTAGCAGCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 215
 QY 414 GAGGTCAACCGCGGGGACACCACTTACGAAGGTAATACTCCAGCTTCTGCTACTAGAAA 473
 Db 216 GAGGTCAACCGCGGGGACACCACTTACGAAGGTAATACTCCAGCTTCTGCTACTAGAAA 275
 QY 474 GTCAGAGAGACTGGCTCCCTCACCTGC 500
 Db 276 GTCAGAGAGACTGGCTCCCTCACCTGC 302

RESULT 2

Z18494/c
 LOCUS 218494 261 bp mRNA linear EST 12-FEB-1993
 DEFINITION ATTS0870 Gif-SeedA+B Arabidopsis thaliana cDNA clone YAYB19T3, mRNA sequence.

ACCESSION Z18494

VERSION Z18494.1 GI:17613

KEYWORDS EST

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 261)

REFERENCE

AUTHORS CNRS.

TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program

JOURNAL Unpublished (1996)

COMMENT Contact: Berthomieu P., Guerrier D., Giraudat J.

Genetique Moleculaire d'Arabidopsis

ISV - UPR40, CNRS

Avenue de la Terrasse, 1198 Gif-sur-Yvette Cedex, France

Email: Giraudat@cnrs-gif.fr.

FEATURES

source

1..261

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/ecotype="Columbia"

/db_xref="taxon:3702"

/clone="YAYB19T3"

/clone_lib="Gif-SeedA+B"

/notes="Vector: Lambda ZAPII; tissues=Green siliques of

A.thaliana ecotype columbia; clone.library=Gif-SeedA+B;

Cloning vector: Lambda ZAPII non-oriented."

ORIGIN

Query Match 33.1%; Score 165.6; DB 7; Length 261;
 Best Local Similarity 93.7%; Pred. No. 7.2e-40;
 Matches 194; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY 294 CTTTCGCGCAGTGAATATGAAGAAGATGAAAAGATTGGTTTGACCGGGAGAACCAATTGA 353

Db 236 CGTGTCTCAGTGAATATGAAGAAGATGAAAAGATTGGTTTGACCGGGAGAACCA--TTA 179

QY 354 CACCAGATCCCTAGCAGCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 413

Db 178 CACCAGATCCCTAGCAGCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 119

QY 414 GAGGTCAACCGCGGGGACACCACTTACGAAGGTAATACTCCAGCTTCTGCTACTAGAAA 473

Db 118 GAGGTCAAGCC-GGGGACACCACTTACGAAGGTAATACTCCAGCTTCTGCTACTAGNAA 60

QY 474 GTCAGAGAGACTGGCTCCCTCACCTGC 500

Db 59 GTCAGAGAGACTGGCTCCCTCACCTGC 33

RESULT 3

B10852

LOCUS B10852 934 bp DNA linear GSS 14-MAY-1997
 DEFINITION F3E13-T7 IGF Arabidopsis thaliana genomic clone F3E13, genomic survey sequence.

ACCESSION

VERSION B10852

KEYWORDS GSS

SOURCE B10852.1 GI:2091974

ORGANISM

Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 934)

REFERENCE

AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.

TITLE BAC End Sequences at ATGC

JOURNAL Unpublished (1997)

COMMENT Contact: Ecker J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenome.bio.upenn.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 129

High quality sequence stop: 183.

FEATURES

source

1..934

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/ecotype="Columbia"

/db_xref="taxon:3702"

/clone="F3E13"

/sex="hermaphrodite"

/clone_lib="IGF"

/note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;

Produced by Thomas Altman"

ORIGIN

Query Match 21.6%; Score 108; DB 8; Length 934;

Best Local Similarity 86.7%; Pred. No. 7e-22; Indels 0; Gaps 0;

Matches 1117; Conservative 0; Mismatches 18;

QY 294 CTTTCGCGCAGTGAATATGAAGAAGATGAAAAGATTGGTTTGACCGGGAGAACCAATTGA 353

Db 429 CGTGTCTCAGTGAATATGAAGAAGATGAAAAGATTGGTTTGACCGGGAGAACCAATTGA 488

QY 354 CACCAGATCCCTAGCAGCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 413

Db 489 CACCAAAATCCCTAGCANCTTCAATTCCTGCTCAGTTGAACAAGAAACCCCTGGTTTGAG 548

QY 414 GAGGTCAACCGCGGG 428

Db 549 GAAGTCAACCCGNGG 563

RESULT 4

CNS00N7X

LOCUS CNS00N7X

DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC F3E13 of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

ACCESSION AL081579

VERSION AL081579.1 GI:5282719

KEYWORDS GSS

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 256)

REFERENCE

AUTHORS Salanoubat, M., Choiane, N., Artiguenave, F., Brottier, P., Wincker, P., Samson, D., Surin, W., Weissenbach, J. and Quetier, F.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 256)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 FEATURES source
 1..256
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="F3E13"
 /clone_lib="IGF"
 /note="end : T7"
 ORIGIN
 Query Match 20.2%; Score 101; DB 9; Length 256;
 Best Local Similarity 100.0%; Pred. No. 7.1e-20;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 204 TCACATTTCCCTCGCGTTTCTTTTTCACACTCTCGAATCGGAATTCGGAATTCACCAAGCTC 263
 Db 1 TCACATTTCCCTCGCGTTTCTTTTTCACACTCTCGAATCGGAATTCGGAATTCACCAAGCTC 60
 QY 264 CGTCAAGCGATAGATTCTGCAATTCACACTTTTCGCGCGAG 304
 Db 61 CGTCAAGCGATAGATTCTGCAATTCACACTTTTCGCGCGAG 101
 RESULT 5
 LOCUS AQ956407/c 719 bp DNA linear GSS 28-JAN-2000
 DEFINITION LERA980TF LERA Arabidopsis thaliana genomic clone LERA980, genomic survey sequence.
 ACCESSION AQ956407
 VERSION AQ956407.1 GI:6784108
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsia.
 REFERENCE 1 (bases 1 to 719)
 AUTHORS Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T., Feldblum, T., Liang, F., Creasy, T. and Fraser, C.M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
 JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tcdb/at/at.html>
 Seq primer: TF
 Class: Shotgun.
 FEATURES source
 1..719
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERA980"
 /clone_lib="LERA"
 /note="Organ: Leaf; Vector: pHOS1. Total genomic DNA was sheared to 0.9-1 Kbp before ligation."
 ORIGIN

Query Match 20.0%; Score 99.8; DB 8; Length 719;
 Best Local Similarity 98.1%; Pred. No. 2.2e-19;
 Matches 101; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 398 AAACCCCTGTTTGGAGGAGTCAAGCGGGGACACCATCTTACGAAGGTAATAACTCCAG 457
 Db 719 AACCCCTGTTTGGAGGAGTCAAGCGGGGACACCATCTTACGAAGGTAATAACTCCAG 660
 QY 458 CTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACCTGC 500
 Db 659 CTCTGCTACTAGAAAGTCAGAGAGACTGGCTCCCTCACCTGC 617
 RESULT 6
 LOCUS BH981035/c 709 bp DNA linear GSS 07-OCT-2002
 DEFINITION odg49f03.b1 B.Oleracea002 Brassica oleracea genomic, genomic survey sequence.
 ACCESSION BH981035
 VERSION BH981035.1 GI:23511755
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 709)
 AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 TITLE Whole genome shotgun reads from Brassica oleracea
 JOURNAL Unpublished (2002)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: odg49 row: f column: 03
 Seq primer: -21UPPOT forward
 Class: shotgun
 High quality sequence start: 96
 High quality sequence stop: 551.
 FEATURES source
 1..709
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.Oleracea002"
 /note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T0100DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."
 ORIGIN
 Query Match 17.7%; Score 88.4; DB 8; Length 709;
 Best Local Similarity 71.0%; Pred. No. 7.5e-16;
 Matches 132; Conservative 0; Mismatches 51; Indels 3; Gaps 1;
 QY 306 GAATATGAGAAAGATGAAAGATTGGTTTGACGGGAGAACCATTTACACAGATCCCT 365
 Db 212 GATGAAGAAAGATGAGAAAGAGGTTTCAAGGGAAGGACCATTTGCACCCAGATC- 156
 QY 366 AGCAGCTTCAATTCCTGCCCTCAGTTGACAGAAACCCCTGTTTGGAGGAGGTCAAGCCG 425
 Db 155 TTCAGCGGTAGCTTCTGCGGTCAACAGTGTAGAAACCACTCTTGGAGGAGGTCAACCCG 96
 QY 426 GGGGACACCATCTTACGAAGGTAAATACTCCAGCTTCTGTCTACTAGAAAGTCAGAGAGACT 485
 Db 95 GGGAAACCCGTCTACGAGCCCTAATACTCCAGTCTTCTAGTACCAGAAAGTCTGGGAGAT 36
 QY 486 GGCTCC 491

```

Db          35  TCCACC 30

RESULT 7
BZ465288/c
LOCUS      BOMHE277F BO.1.6.2 KB tot Brassica oleracea genomic clone BOMHE277,
DEFINITION genomic survey sequence.
ACCESSION BZ465288
VERSION    BZ465288.1 GI:26753145
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOMHE277R
Contact: Chris Town
TIGR
7112 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES             source
    Location/Qualifiers
        1..512
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOMHE27"
            /clone_lib="BO.2.3 KB"
            /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match      13.9%; Score 69.4; DB 8; Length 512;
Best Local Similarity 65.6%; Pred. No. 5.2e-10;
Matches 141; Conservative 0; Mismatches 56; Indels 18; Gaps 2;

QY 303 AGTGAATATCAAGAAAGATGAAAGATTGGTTTGACGGGGAGAACCAATTACACCAAGATC 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 AGATGAGATGAAGAAAGAGAGAGAGATGTTTCACGGGGAGGACTATTTCGACTAGGTT 382

QY 363 CTAGCAGCTTCAATTCTCTGCTCAGTTGAAACAAGAAACCCCTG-----GTTTGAGGAG 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 TGTAGCTGTAGCTTCTCTGCTCAGCGGAGAAAGATATTCTGGTTTGAGTTTGAGGAG 322

QY 417 GTCAGCCGGGGGACACCATCTAC-----GAAAGTAATAACTCCAGCTTCTGC 464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 GTCAGCCGGGGGACATCGTGACTACTTCTAGGAAGAGTAGCAACCAACTTCTAG 262

QY 465 TACTAGAACTCAGAGACTGGCTCCCTCACCTG 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TACCAGCAAGTCGGAGAACGGAATCCTTCACCCG 227

RESULT 9
BZ465288/c
LOCUS      BOMHE277F BO.2.3 KB Brassica oleracea genomic clone BOMHE27,
DEFINITION genomic survey sequence.
ACCESSION BH707844
VERSION    BH707844.1 GI:18793338
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
BAC End Sequences at ATGC
Unpublished (1997)
Other GSSs: F18C19-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA

BZ465288
LOCUS      BOMHE277F BO.1.6.2 KB tot Brassica oleracea genomic clone BOMHE27,
DEFINITION genomic survey sequence.
ACCESSION BZ465288
VERSION    BZ465288.1 GI:26753145
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Contact: Chris Town
TIGR
7112 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES             source
    Location/Qualifiers
        1..684
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOMHE27"
            /clone_lib="BO.1.6.2 KB tot"
            /note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
            total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match      14.5%; Score 72.6; DB 8; Length 684;
Best Local Similarity 66.5%; Pred. No. 5.8e-11;
Matches 143; Conservative 0; Mismatches 54; Indels 18; Gaps 2;

QY 303 AGTGAATATCAAGAAAGATGAAAGATTGGTTTGACGGGGAGAACCAATTACACCAAGATC 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 AGATGAGATGAAGAAAGAGAGAGATGTTTCACGGGGAGGACTATTTCGACTAGGTT 415

QY 363 CTAGCAGCTTCAATTCTCTGCTCAGTTGAAACAAGAAACCCCTG-----GTTTGAGGAG 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 TGTAGCTGTAGCTTCTCTGCTCAGCGGAGAAAGATATTCTGGTTTGAGTTTGAGGAG 355

QY 417 GTCAGCCGGGGGACACCATCTAC-----GAAAGTAATAACTCCAGCTTCTGC 464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 GTCAGCCGGGGGACATCGTCTACTTCTAGCAAGAGCTAGCAACCAACTTCTAG 295

QY 465 TACTAGAACTCAGAGACTGGCTCCCTCACCTG 499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 TACCAGCAAGTCGGAGAACGGAATCCTTCACCCG 260

RESULT 8
BH707844/c
LOCUS      BOMHE277F BO.2.3 KB Brassica oleracea genomic clone BOMHE27,
DEFINITION genomic survey sequence.
ACCESSION BH707844
VERSION    BH707844.1 GI:18793338
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
BAC End Sequences at ATGC
Unpublished (1997)
Other GSSs: F18C19-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA

```

19104

Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@genome.bio.upenn.edu
 Seq primer: Sp6
 Class: BAC ends
 High quality sequence start: 152
 High quality sequence stop: 206.
 Location/Qualifiers
 1. .926
 /organism="Arabidopsis: thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="F18C19"
 /sex="hermaphrodite"
 /clone_lib="IGF"
 /notes="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
 Produced by Thomas Altmann"

FEATURES
source

ORIGIN

Query Match 11.9%; Score 59.4; DB 8; Length 926;
 Best Local Similarity 84.9%; Pred. No. 7.7e-07;
 Matches 79; Conservative 0; Mismatches 11; Indels 3; Gaps 1;
 QY 214 CTCGCGTTTTTCTTTTCCACA---CTCTCGGAATCGGAATTTCCACCAAGTCGGTCAAG 270
 Db 113 CCGCGTTTTTCTTTTCCACACTCTCCGGAATCGGTAATTTCCACCAAGTCGGTCAAG 172
 QY 271 CGATAGATTCGACATTTACACACTTTCGCGCA 303
 Db 173 CGATAGATTCGACATTTACACACTTTCGCGCA 205

RESULT 10

CR464122 273 bp mRNA linear EST 01-JUL-2004
 LOCUS CR464122 Rat pBluescript Lion Rattus norvegicus cDNA clone
 DEFINITION CR464122
 ACCESSION CR464122.1 GI:49596471
 VERSION CR464122.1
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 273)
 Reference
 Henrich, J., Hermanns, J., Kranz, H., Loebbert, R., Schlueter, T.,
 Schuetz, D., Weindel, M., Heil, O., Ebert, L., Neubert, P., Peters, M.,
 Radloff, U., Schneider, D. and Korn, B.
 Rat ArrayTAG cDNA
 Unpublished (2004)
 Title
 Rat ArrayTAG cDNA
 Journal
 Contact: Inge Arlt
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Heubnerweg 6, D-14059 Berlin, Germany
 Email: www.rzpd.de
 RZPD: LIONp463C08360.
 RZPDLIB;
 Rat ArrayTAG cDNA
 http://www.rzpd.de/cgi-bin/products/showlib.pl.cgi?response?libNo=463 Contact: Inge Arlt
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Heubnerweg 6, D-14059 Berlin, Germany
 Tel: +49 30 32639 100
 Fax: +49 30 32639 111
 www.rzpd.de
 This clone is available royalty-free from RZPD;
 contact RZPD (clone@rzpd.de) for further information. Seq primer:
 RP: CAGGAACACGCTATGAC.
 Location/Qualifiers
 1. .273
 /organism="Rattus norvegicus"

FEATURES
source

/mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="LIONp463C08360"
 /lab_host="DH108"
 /clone_lib="Rat pBluescript Lion"

ORIGIN

Query Match 7.6%; Score 38.2; DB 7; Length 273;
 Best Local Similarity 55.7%; Pred. No. 2;
 Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
 QY 25 GTAATCGTAAATCAAAATCGCTCGCTCGAGAAACCGCGTCTAGAGAGGAAAGTAC 84
 Db 28 GTAATTGTATAAACAACAAGAACTTCTCTAGGAACAACAGAAATTAANAAGCAAGGCC 87
 QY 85 CGTACATAATCTCGGAAACCAATTCGCTCTTCTTCAAACTCAGTTTTCGGAACCCCA 144
 Db 88 TCTGCATCCCAAGTAACACACAGTCGCTGTTTCCACACAGACCCCGCAGATCACC 147
 QY 145 AACACCGCGAG 155
 Db 148 AACCTGCTGAG 158

RESULT 11

BI503269 541 bp mRNA linear EST 08-APR-2002
 LOCUS BI503269/c
 DEFINITION BI503269 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170008B20B03 5', mRNA sequence.
 ACCESSION BI503269
 VERSION BI503269.1 GI:15353643
 KEYWORDS EST.
 SOURCE Apis mellifera (honey bee)
 ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
 1 (bases 1 to 541)
 Reference
 Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L.,
 Pardini, J., Robertson, H.M., Soares, B. and Robinson, G.E.
 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
 Genome Res. 12 (4), 555-566 (2002)
 JOURNAL MEDLINE
 PUBMED 11932240
 COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499
 Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
 REPEAT IN THE SEQUENCE
 Simple repeat STRAND (+) ELEMENT (CAT)n LOCATION [365,403].
 PCR Primers
 FORWARD: TAATACGACTCTACTATAGG
 BACKWARD: ATTAACCTCTACTAAG
 Plate: BB170008B20 row: B column: 03
 Seq primer: AGCGATAACAATTTACACAGGA
 High quality sequence stop: 541.
 Location/Qualifiers
 1. .541
 /organism="Apis mellifera"
 /mol_type="mRNA"
 /strain="mixed strains of European bees, predominantly A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="BB170008B20B03"
 /sex="female"

FEATURES
source

```
/tissue type="brain"
/dev stage="adult worker honey bee"
/lab_host="DH10B"
/clone lib="Bee Brain Normalized/Subtracted Library, BB17"
/notes="Organ: brain; Vector: pTT3-Pac; Site_1: EcoRI;
Site_2: NotI; this BB17 cDNA library was generated by
subtraction of the BB16 library with 4000 previously
sequenced clones. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
```

ORIGIN

```
Query Match 7.6%; Score 38.2; DB 4; Length 541;
Best Local Similarity 50.3%; Pred. No. 2.4; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 188 AATTGTGCTGGAATCTCAAAATTTCCCTCGCGTTTCTTTTCTTTCACACTCTCGGAATCGGA 247
|||||
Db 289 AATTCCATCGTTGTTGCAATTTTCCCGGATATTTGTATATATCTCGATCGATAC 230
|||||

QY 248 AATTCCACAGCTCCGTCAGCGATAGATTTCTGCAATTCATACACACTTTCCGGCGAGTGA 307
|||||
Db 229 GATTTTCACCTTCGTGCAATTTGCAATTTGTTGCGCGAATTTGAAATTCGCTTCTATTTC 170
|||||

QY 308 ATATGAAGAAGATGAAGATTTGTTTCACGGGAGAACCATTTACACAGATCCCTAG 367
|||||
Db 169 TAATGAAAAAGGAAAAAATCGATTACGGAGAAATTTTTCATTTGTTATTCAG 110
|||||

QY 368 CAGCTTC 374
Db 109 AATGTC 103
```

RESULT 12

```
AQ807478
LOCUS
DEFINITION
HS 3207 B1_D03_T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens_genomic clone Plate=3207 Col=5 Row=H, genomic survey
sequence.
```

ACCESSION

```
AQ807478
AQ807478.1 GI:5726720
```

VERSION

```
AQ807478
GSS.
```

KEYWORDS

```
GSS.
```

SOURCE

```
Homo sapiens (human)
```

ORGANISM

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 531)
```

REFERENCE

```
AQ807478
531 bp DNA linear GSS 10-AUG-1999
sapiens_genomic clone Plate=3207 Col=5 Row=H, genomic survey
sequence.
```

AUTHORS

```
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
```

JOURNAL

```
99380589
MEDLINE
PUBMED
```

COMMENT

```
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3207 row: H column: 5
Seq primer: T7
Class: BAC ends
High quality sequence stop: 531.
```

FEATURES

```
Location/Qualifiers
```

```
1..531
```

source

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3207 Col=5 Row=H"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pBelO8A11; BAC Clones in
E-Coli DH10B"
```

ORIGIN

```
Query Match 7.4%; Score 37.2; DB 8; Length 531;
Best Local Similarity 65.9%; Pred. No. 4.9; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 308 ATATGAAGAAGATGAAGATTTGTTTCACGGGAGAACCATTTACACAGATCCCTAG 367
|||||
Db 273 AATTACATAGGTTCAAGATTTGTTTGACCGGGTGTTACCATTTACATACCTGCAAA 332
|||||

QY 368 CAGCTTCAATTCCTGCCTCAGT 389
|||||
Db 333 AAACITGCTTCCCAACCCCAAT 354
|||||
```

RESULT 13

```
CB404139/c
```

LOCUS

```
CB404139.1
```

DEFINITION

```
EST.
```

ACCESSION

```
CB404139.1
```

VERSION

```
GI:30745866
```

KEYWORDS

```
Caenorhabditis elegans
```

SOURCE

```
Caenorhabditis elegans
```

ORGANISM

```
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
```

REFERENCE

```
1 (bases 1 to 553)
```

AUTHORS

```
Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M.,
Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T.,
Hudson, J.R., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S.,
Endress, G.A., Jenna, S., Chevet, E., Papasotiropoulos, V.,
Tollas, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H.,
Doucette-Stamm, L., Hill, D.E. and Vidal, M.
```

TITLE

```
C. elegans ORFeome version 1.1: experimental verification of the
genome annotation and resource for proteome-scale protein
expression
```

JOURNAL

```
Nat. Genet. (2003) In press
```

COMMENT

```
Contact: Vidal M
```

```
Marc Vidal Laboratory
```

```
Dana Farber Cancer Institute
```

```
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
```

```
Tel: 617 632 5180
```

```
Fax: 617 632 5739
```

```
Email: Marc.Vidal@dfci.harvard.edu
```

```
Sequence tag of Gateway entry clones. The primers used were
```

```
designed on the predicted protein encoding ORF. C. elegans ORFeome
```

```
cloning project : Contact david_hillefci.harvard.edu or
```

```
marc_vidal@dfci.harvard.edu
```

```
POLYA=No.
```

FEATURES

```
Location/Qualifiers
```

```
1..553
```

source

```
/organism="Caenorhabditis elegans"
```

```
/mol_type="mRNA"
```

```
/strain="N2"
```

```
/db_xref="taxon:6239"
```

```
/sex="Hermaphrodite and male"
```

```
/tissue type="whole animal"
```

```
/dev stage="mixed stage"
```

```
/clone_lib="AD-wrmcDNA"
```

```
/note="The AD-wrmcDNA library was generated with poly(A)+
```

```
RNA isolated from both hermaphrodite and male N2 worms of
```

```
all larval stages, embryos, adults and dauers and the
```

```
subsequent generation of cDNAs by poly(A) priming. The
```

```
cDNAs were cloned into pPC86"
```

ORIGIN

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2534.52 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929C-2_COPY_3000_3500

Perfect score: 501

Sequence: 1 agtcataacactccagtcacat.....acacaaggttgagttcgatcgca 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	6554	8	AF213627 Arabidops
2	501	100.0	6571	6	AX063583 Sequence
3	224	44.7	10329	6	AX063582 Sequence
4	224	44.7	10329	8	AF213628 Arabidops
5	224	44.7	120965	8	AC026875 Genomic s
6	180	35.9	6995	8	AFH552638
7	70.2	14.0	4027	8	AK068790 Oryza sat
8	63.4	12.7	417	6	AX063614 Sequence
9	57.6	11.5	515	6	AX063611 Sequence
10	42.6	8.5	2000	6	AX655393 Sequence
11	41	8.2	125020	9	AF429315 Homo sapi
12	40.4	8.1	209378	2	AC151545 Callichri
13	39.8	7.9	193969	5	BX294103 Zebrafish
14	39.4	7.9	154492	2	AC149661 Bos tauru
15	39.2	7.8	10303	1	AE013102 Thermoana
16	38.2	7.6	198057	2	AC126072 Rattus no
17	37.4	7.5	2000	6	AX655393 Sequence
18	37.2	7.4	135752	9	AC073465 Homo sapi
19	37.2	7.4	144007	2	AC141743 Apis mell

20	37.2	7.4	208729	2	AC012052	AC012052 Homo sapi
21	37.2	7.4	215177	2	AC011667	AC011667 Homo sapi
22	37.2	7.4	253420	10	AC100893	AC100893 Mus muscu
23	37	7.4	212079	2	AC134605	AC134605 Mus muscu
24	36.8	7.3	1141	6	AX083744	AX083744 Sequence
25	36.8	7.3	42114	3	U50193	U50193 Caenorhabdi
26	36.8	7.3	96975	9	AC004975	AC004975 Homo sapi
27	36.8	7.3	216491	2	AC128110	AC128110 Rattus no
28	36.8	7.3	231157	2	AC111347	AC111347 Rattus no
29	36.6	7.3	156150	5	BX322540	BX322540 Zebrafish
30	36.6	7.3	181034	2	AC135486	AC135486 Rattus no
31	36.6	7.3	190724	2	AC078877	AC078877 Homo sapi
32	36.4	7.3	148235	9	AC146372	AC146372 Pan trogl
33	36.4	7.3	162995	10	AL663032	AL663032 Mouse DNA
34	36.4	7.3	169102	9	AL359997	AL359997 Human DNA
35	36.4	7.3	199271	2	AC142243	AC142243 Mus muscu
36	36.4	7.3	207495	10	AL772393	AL772393 Mouse DNA
37	36.4	7.3	263608	2	AL772254	AL772254 Mus muscu
38	36.2	7.2	110000	8	CR382132	Continuation (26 o
39	36.2	7.2	151902	2	AC015668	AC015668 Homo sapi
40	36.2	7.2	154713	2	AC010958	AC010958 Homo sapi
41	36.2	7.2	166507	2	AC151442	AC151442 Atelalex
42	36.2	7.2	169813	9	AC114341	AC114341 Homo sapi
43	36	7.2	16460	3	AC024807	AC024807 Casnothab
44	36	7.2	110000	8	CR382123	Continuation (14 o
45	36	7.2	153054	5	AC145789	AC145789 Xenopus t

ALIGNMENTS

RESULT 1
AF213627
LOCUS Arabidopsis thaliana MOM mRNA, complete cds.
DEFINITION Arabidopsis thaliana MOM mRNA, complete cds.
ACCESSION AF213627.1 GI:8132767
VERSION AF213627.1
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE 1 (bases 1 to 6554)
Amedeo,P., Habu,Y., Afsar,K., Scheid,O.M. and Paszkowski,J.
Disruption of the plant gene MOM releases transcriptional silencing of methylated genes
NATURE 405 (6783), 203-206 (2000)
20279299
PUBMED 10821279
REFERENCE 2 (bases 1 to 6554)
Amedeo,P., Habu,Y., Afsar,K., Mittelsten Scheid,O. and Paszkowski,J.
Direct Submission
Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box 2543, Basel 4002, Switzerland
Location/Qualifiers
1. .6554
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Zhuerich"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATEAT1 and phyA"
310. .6315
/function="Involved in the maintenance of transcriptional gene silencing"
/codon_start=1
/product="MOM"
/protein_id="AAF73380.1"
/db_xref="GI:8132768"
/translation="MKDKSIGLGTGTYITRSLAIPASVEQETGLRRSSRGTRST
KVITPASATRKGERLAPSVKSGGIVKQNSTPSSLRNRNKGTEVLSQSKGSDN

SIRKGDTS PDI EQRKDS VBESTD KIKP IMSARSYRAL FRGKUKESBAL VDA SPNEBEL
VWGSRRIPAGNDVDQCPDPADAGSKRLPVDETS LDKGTDFPLKSVTETEKIY
LDASPIVETDSDVSGSPENLTKQJOKGDKTQSPANAAESKTPVGSSTVETEKIYQ
KFCQSDNTLECPANAAESKRLPVGETSLEKDTDFPLKSTETGOMVLASPIVETRDS
VQSPSTNLETKQLLVSKTGLIEVILPLRKRDPTAEIDCATVANGDVMSSDGS
VIPSQCKNDNRPEMCKNTCKRKQKVDGQNRSCVSCSIVQVPEBQDNTQDMKTEGP
VTSREYBENGQHQGKSSDPKSYSSVYBYPVVPVLQVLOEYQCOYLFSKLSLSL
SKDILGALETSVRKTCDPHYVMDASLQJLTKNLELIEILDVEIKASGKJLHJLQK
NHLIKKNGKALVAVPQTSQTIKJLRADAFILFGSLNRPGRKSYHEGIVKISCSRSAL
MFKNECSOVLILLFETACQSIKJLRADAFILFGSLNRPGRKSYHEGIVKISCSRSAL
KIPRLYSVTEBEKALLILARQNMQRKQAVENLRNLSITALLAMGASVLPDKLDPHFS
ETPDVSGVSEFOSIMDGVITEFSSILSSGKBNENVKILLALLMGAQYQTS SDSLTFGS
EDHILKLDSESPNLFMSKLGGKPNMKYPSDTPORNKRVOYFEGSEASPTKGTDCGN
AKKRSKASDDVTDPRVDDKSGKGMKAGKMGALSPKVIITLQSCSKSGTGDGL
DGNDAGLYSMGSHISGIPEDMDLASQDMQKIPDESGORRHVLKPMKALCQVILHDL
CATSMVGNLFEYVYENHRIYEEPAFTTQAFQALGQVIAALLVQKILSHKSELVRANSE
LAFKRSRVEDVYIYILSCMSKLFLEHTQJGFCFTGMSKQSVSTKLNVNLSGASAT
VRDEKINTSMRNSDESECKTEKCSHYSTATDTEKTIJSQKKYKQVQKJVEH
BEKWEILLNMVADKQKULETSKSVAAVYITCSRSTTSQDGLKJLHJNKRKFDEIK
SEKNECLKLSQMEHVAKKLAEADCAWLNIRKSWAKLKVCVPLQSGNKHKEFGSSRN
ISQAPDVQI GNNANVEATYADTCHASKVQVPAENLTGTMSSGSGSTQVHEVDVDR
NDETMDVALSRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEGHEQNCDRIT
SAAEDSDVSRPEVMSQENLASVPSNRAESQYDQIITENNRRTSHVGFDQNPIDQ
NRDCSDLSQETPDELMANVHLASVYETNRGAESEYQGDQICPMFSSLGTQNPIDQ
TESNLEBAIPQASGESVETIDFPAASHQGVQCVLPSSTGQTPQAPANI EGQNI
NTSABPHVAGPDVAGSBDVAVIQETMGAQACSLPSGSGVQSDGLJANEQONTVV
AQLPTDCSDAVVTGSPSVSDCAQDAPMELSSPGNHDPDTANIEGLNTNVAEPHIS
GSDACEMBISPEFGOVERSEANLIFHEGGEVHASQVTAI VPSLLNNGTQLQVAPVQ
IPFPNDVSLRQELTKQSNEHASIYPEILLIPADQCEFPALNVEG

ORIGIN

Query Match	100.0%;	Score 501;	DB 8;	Length 6554;
Best Local Similarity	100.0%;	Prod. No. 1.7e-133;		
Matches 501; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	AGTCATAA	CACTCCAGTCATCATGTAAATCTTCTGGTACAGATGGTACATTCGATGGAAA	60
Db	3000	AGTCATAA	CACTCCAGTCATCATGTAAATCTTCTGGTACAGATGGTACATTCGATGGAAA	3059
Qy	61	TGATGCTTTTGGCTTGTATTATGGCAGCCATATCTCTGGAATCCAGAGGATATGTT	120	
Db	3060	TGATGCTTTTGGCTTGTATTCTATGGGCAGCCATATCTCTGGAATCCAGAGGATATGTT	3119	
Qy	121	AGCTAGTCAAGATTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT	180	
Db	3120	AGCTAGTCAAGATTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT	3179	
Qy	181	AAAGCCGAAGATGGCAAAATCTTTGCCAAGTTTTTGCATCTTTTCAGATGCTTTGCAACAAGCAT	240	
Db	3180	AAAGCCGAAGATGGCAAAATCTTTGCCAAGTTTTTGCATCTTTTCAGATGCTTTGCAACAAGCAT	3239	
Qy	241	GGTCGGAAATTTTCTCGAATATGTTATTTGAAATCACCGAATCTACGAAGAGCCAGCCAC	300	
Db	3240	GGTCGGAAATTTTCTCGAATATGTTATTTGAAATCACCGAATCTACGAAGAGCCAGCCAC	3299	
Qy	301	TACTTTTTCAGGCATTCCAGATAGCCCTGAGTTGGATTGCAGCCTTGTTCGTAAGCAAAAT	360	
Db	3300	TACTTTTTCAGGCATTCCAGATAGCCCTGAGTTGGATTGCAGCCTTGTTCGTAAGCAAAAT	3359	
Qy	361	TCTTAGGCCACAAAGAAATCTCTGTGTCCTGCAAAATTTGAAATTAGCTTTTCAATGCTCTAG	420	
Db	3360	TCTTAGGCCACAAAGAAATCTCTGTGTCCTGCAAAATTTGAAATTAGCTTTTCAATGCTCTAG	3419	
Qy	421	AGTAGAGGTGGATTATATTTTATTCGATATGTCCTGCAATCAAGAGTCTGTTCTCTGGAGCA	480	
Db	3420	AGTAGAGGTGGATTATATTTTATTCGATATGTCCTGCAATCAAGAGTCTGTTCTCTGGAGCA	3479	
Qy	481	TACACAAGGTTTGCAGTTCGA	501	

Db 3480 TACACAGGTTTCGAGTTTCA 3500

RESULT 2
AX063583
LOCUS
DEFINITION
Sequence 2 from Patent WO0100801.
AX063583
ACCESSION
Version
AX063583.1 GI:12541307
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
Habu, Y., mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
Gene involved in epigenetic gene silencing
Patent: WO 0100801-A 2 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

FEATURES
source
1..6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310..6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
trnalaatons="MKKDEKIGLTGRTIVTSLAASIPASVEOTETGLRBSRGTSTP
KVTPASATKSRRLAPSNASVSKSGGIVKNSTPSLRSNRGKTEVSLQSSKGTSD
STRKGDTPDIQRKDSVBESTDKIKPIMSARSYRALFRGLKESALVDASNEBEL
VWVGSRRIIPAGNDVVGKTDICPPADGASKRLPVDETSLDGKTDLPKSVTEKEIV
LDASPIVETGGDSVIGSBNLETKIQDGKDCSPANAESKTLPGVETSLTEKIV
KFQDNDTCLPPANAESKRLPVGETSLEKTDPLKSTTETGKMLYASPIVETRSSD
VICSPNTNLTAKSLVSKTGLETDIVLPLKRRDRTABIELDACAIVANGDDHVMTSSD
VIPSCKNDNRPENCNTCKRKQVNGDCQNSVCSCIQPVVEESDNTVQDMKETGP
VTSREYENGQIQHGKSDSPKFYSVUYPEYWPVQLSDIQLEQYCTLFSSKLSLASL
SKLDGLAETLSNVRKTDCHPVMASLQKMLTLEHEILDLVEIKASGKLHLDDK
MLTHIKKNGLAIVFYQATQTPGLLIGNILEDFVGQFGPKSYEHGIVSEKSNKAIN
NFKNSQCCVLLLETSQTIKLLRADAPILFGLSNPDSVDHYKVEIKIESCSERT
KIFRLYSVTEBKALILARQNRQKNAVENLRSLTHALLMGASVLFQKLDHFHFS
ETPDGVSFQSTMDGVIHFESILSSKGEEENEVKCLLEAKHAQTVYSSDSTLFG
EDHIKLSDRESPIVFSKLJGDNPMWKYPSDTPORNRKRVQYFEGSEAPKPTGDGN
AKRKKASDDVTPRVDTPVDDEKASKGHDHGALESFKVITLQSSCKSGTGTGL
DNDGAPFLYSMGSHISGIPEDMLASQMGKIPDESORRLHTVLKPKWAKLCQVHLHSD
ACTSMGVNFEVYIENRIYEEPATTFQAFIALSWTALLVQVLSHKESSLVRANSE
LAFPKRESVDYIYISLRCKMSLFLHTQGLQDFGTNSKQVLSKLVNLSGLAT
VREKNTKMSRNSDEECMTSEKRSYSTARIDEKTIKTKKQVKKQVQLVORH
EKKMELLNNAADKKOLETSFSAEAAVIRITCSRTSTQVGLDKLIDHNTERKFDFTI
SENECLKSLQEWHEVAKKLADEACWNRIKSWAAKLAVCVPIQSGNNKHFSGSN
ISQNAEDVQICNNAYEATYADNTCMASKVNOYPEAETLGTWGGASTQCVHEMDVVR
NDETMDVSLRQLTKSQNEHASTVPEILIPADCCQEFALNLVHLSDDQCDRITL
SAASDESDSVRSPEVQSLNLSASPEFLSNREELVTTENRRTSHVGFDTNLDLQOQ
NREDCSLDQRIPELPAVQHLASVVEYTRGAESDQYQDICIEMPSLACGQVDPDPA
TESNELEAPEDQASGETVETTDFAASHQGDQVTPCLSSPTGQNPAPANTEGQNI
NTSAAEPHAEVAGSDGYAVIQETWGAQDACSPLSGSVGTQSLDGLANTEGQNVTV
AOIPTDGSDAVTVTSGSPVQDCAQDASPMELSPGNHDPDTAVNI EGLDNTSAEPHLS
GSDCAWMEISEPGQVSTRTANLPHFEGVHSGAGVTALVPSLLNNGTEQIVAVQVPO
IIPFVNDPLFLEHLEKLRSESNKTFEEKKSILKAEERKMAEVQERFRKRFHEVE
AENRTTKIETKDNILVIMNKLNAFLSKCTDKKVSFSGAPKQSIQOLQARAQVSA
LRYNIATFQKASFPALVAPLQLOQSSFPAPGPAPLQPSQKSIQOLQARAQVSA
LNFVACMPGPOPLISNIATPTSTPATNPRLSPHLSNPSRSSSTVPATPTPTS
SVFPOALITYSAVISIOOEOFOPOOSISSGIOSNNEVYVCLSDDE

CDS

ORIGIN

Query Match	100.0%;	Score 501;	DB 6;	Length 6571;
Best Local Similarity	100.0%;	Pred. No. 1.7e-133;		

1 AGTCATAACACTCCAGTCATCATGTAAATCTTCTGGTACAGATGGTACATTGGATGGAAA 60

CDS
PIVADSEGTDTNPSYSLDALAVFPRVLRVLRVHFKGLDKSSNAGYVLLMFYHLYEGKN
RNFESLIRGSLIKMPLKSDRTPLPDPVKSJVLSEBGLDLFLNLSRRHRLESTKG
TYAEMKRSQKQRLDIAVANSYLSIQVFESFVQVHVRLEELTIAKDYKPPSSSEKR
KGSIVFAAIPNATQVHSLLEKLAANPTMRSPFLGKKKSIQEKLSRHVTLIAHRS
HGATVANSQHLNRVPSLTELINDKMAALTARHGVSDGETVWSKNEMPHVLTWT
AEGVTAKEANTLPQVLEGKASRLVDPVPSISGLEPFF"
join(11454..11915,1209..12885,12925..13073,14200..14235,
14754..14777,14903..15370,15466..16353)
/notes="putative elongation factor 1-a
gb|AA13128.1|AF124801.1; similar to ESTs gb|AW398653.1,
emb|Z17990.1, gb|AW220334.1, gb|AW622026.1,
dbj|AV439770.1, gb|T45275.1, and gb|AW267999.1,
gb|T04414.1, gb|AW221780.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.2"
/protein_id="AAF79842.1"
/db_xref="GI:8778823"
/translation="MGKEPHINIVIGHVDSKSTTTHLYIKLGGIDKRVIERPEK
EAAEMKRSFYAWLVDKLAERERGITIDIALWKPEFTTKYCTVIDAPGRDPIKNN
ITQSDACAVLIIDSTTGGFEAGISKQGTREHALLAFTLVKQMICCNKMDATTP
KYSKARDEIIEKVSYLKKVNPDKIPFVPSGEGDNMIERSNLDWYKPTLLE
ALQDINEPKRPSDKPLRLPLQDVYKIGGIVPVRVETGMIPKGMVVPFAPTLTE
VKSVEMHESLLEALPGDNVGNVAVKDLRGYVANSKDDPAKGAANFTSQVIL
MNPQGQNGYAPVLDSHTSHIAVSEILTKIDRSKGKEIEKEPKFLKNGDAGVMK
TPKPMVETFSYPPPLGRFAVRDMRQTVAVGVKSVDDKDPGAKVTKAAVKKGAK
DTSVFTIWMVLVACISCFVSPSCSNWLDRLRWIRYTHLPHPSPICDVC
MIEHNFVLMGKEPHINIVIGHVDSKSTTTHLYIKLGGIDKRVIERPEKAAE
MNRSPFYAWLVDKLAERERGITIDIALWKPEFTTKYCTVIDAPGRDPIKNNITG
SOACAVLIIDSTTGGFEAGISKQGTREHALLAFTLVKQMICCNKMDATTPKYSK
ARDEIIEKVSYLKKVNPDKIPFVPSGEGDNMIERSNLDWYKPTLLEALDQ
INEPKRPSDKPLRLPLQDVYKIGGIVPVRVETGMIPKGMVVPFAPTLTEVKS
EMHESLLEALPGDNVGNVAVKDLRGYVANSKDDPAKGAANFTSQVILNHP
GOIENGYPALDCHTHIAVKFSEILTKIDRSKGKEIEKEPKFLKNGDAGVMKWTPTK
PMVETFSYPPPLGRFAVRDMRQTVAVGVKSVDDKDPGAKVTKAAVKKGAK"
complement(join(19245..20132,20226..20687))
/notes="putative elongation factor 1-a
gb|AA13128.1|AF124801.1; similar to ESTs gb|AW034714.1,
gb|A1812185.1, emb|Z17990.1, gb|AW267999.1, gb|N96009.1,
dbj|AV439770.1, and gb|T45268.1"
/codon_start=1
/evidence=not experimental
/product="T6D22.3"
/protein_id="AAF79847.1"
/db_xref="GI:8778848"
/translation="MGKEPHINIVIGHVDSKSTTTHLYIKLGGIDKRVIERPEK
EAAEMKRSFYAWLVDKLAERERGITIDIALWKPEFTTKYCTVIDAPGRDPIKNN
ITQSDACAVLIIDSTTGGFEAGISKQGTREHALLAFTLVKQMICCNKMDATTP
KYSKARDEIIEKVSYLKKVNPDKIPFVPSGEGDNMIERSNLDWYKPTLLE
ALQDINEPKRPSDKPLRLPLQDVYKIGGIVPVRVETGMIPKGMVVPFAPTLTE
VKSVEMHESLLEALPGDNVGNVAVKDLRGYVANSKDDPAKGAANFTSQVIL
MNPQGQNGYAPVLDSHTSHIAVSEILTKIDRSKGKEIEKEPKFLKNGDAGVMK
TPKPMVETFSYPPPLGRFAVRDMRQTVAVGVKSVDDKDPGAKVTKAAVKKGAK"
complement(20787..20857)
/product="tRNA-Lys"
/notes="codon recognized: AAA; putative tRNA-Lys"
/evidence=not experimental
complement(join(21956..22189,22535..22600,22713..22929,
23041..23126))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.4"
/protein_id="AAF79846.1"
/db_xref="GI:8778847"
/translation="MCLNRLPSATPIPRAREERECOSMFGFRNHQLLLSSFLCKRPA
IINWKGSGGSGSGSGPTAAAAAALQKALLQRTVDTITSVDNFTQIVNRSVR
SDPPVNSQSTYNNMRASMQVAAASLLKLVSELKQTAIFSGFASLNDHVEQRIEF
DQAEKNTLRLIARADASANKELSHYSSAQRLLDI"
join(231576..23689,24080..24258,24347..24522,24635..24726)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental

CDS
/product="T6D22.5"
/protein_id="AAF79823.1"
/db_xref="GI:8778824"
/translation="MTGLARLVAPMILLIFIELVKAETVITLTPTFSDKLALSN
KQALONGCSWCSFYSLSDDTTRCLLSPCLLIVQIKKOTAFWPKVCPWCKKCKLIG
NWEDLIGKAMEGDEIEVEGDCGTSTRAVTKVIEHSYPTFMLFYNGEESYKVGKRD
VBSLKAFVVEETKAAKKAQEDKEL"
complement(join(25423..25563,25675..26048,26127..26330,
26585..26783,26941..27842,27936..27991,28208..28413))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translation="MDSVRRKINNTTQSSPSKPSKFSVYRNPLAAAASTANSIRP
SKSVLLPILLSIASFSLISPMAGEKMLTNALTFKISQEAAYVTVKAWQGLVTLFC
IGAMWALSGISLHRAKFAARGETKSFSTKDOFSLSNRQLELLGKKKADQSVSFC
PKSRPAKPAQSLPLVPIHQTLTGSAAHSGGGKLNRSNGSQISSFTSPKQMSG
PSMYLVSPSSVSSNLSQDKAVSSPWSGRSSAKDITTEQLQLLAEIDKITE
SAGKMTPTPTVGSFAMASPTVTRSGASGATSTPLRVRMSPGAKFTTTPPKGE
GPPNPMISLERAIEGFIHLGVYQIEDMDRLQWQSVLLKPLLNKVTSHIQVMOT
ASKLGVNVTVSQVSDLPNTGTATTALPVDRTKGWOPSYSLDEDALLHOLRANLVQAI
DASOKLRTENQFQOQQOQQOQAAALIPWQECVDAISEHRLQLGKMGKGLPLP
RSIPADYTVQIRALAEGTGVNRYNGADTRKNNKWSLPPPTDLSHLLYLFCAP
LHPKMWLHLDPSYTGTAQKNPLFLGLVLPKPRPEKYIAVYGVSPVSTHFGACVL
AVDQKQPPFTALYMDKKVQFTLQGRALWDSMLLICHRIVGYGVVGRMNLGSSALN
ILQVDSDDTD"
complement(join(29418..29491,29579..29620,29858..29942,
30386..30805))
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translation="WVSSKPKPKKARSVVVNVKASGRSKRSGSRTKTSNKNVIVK
KKEPIEISESSSSVEEAIKDEAKNSGVVSKRNGSKSVGIPTKTSKNREDDGG
ADEAKIKFPMNRIRIRIMRDSNAPQIMQDAVLNATKATMFIERFSEAYDSSVKKDK
KETHYKHLSSVSNDRQYEFLLADSVPEKLKAEALBWERGMTAG"

Query Match 44.7%; Score 224; DB 8; Length 120965;
Best Local Similarity 100.0%; Pred. No. 1.5e-53;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTCATAACATCCAGTCATCATGTAAATCTTCTGGTACAGATGGTACATGGATGAAA 60
|||||
DB 62689 AGTCATAACATCCAGTCATCATGTAAATCTTCTGGTACAGATGGTACATGGATGAAA 62630
|||||
QY 61 TGATGCTTTTGGCTTGTATTCTATGGGACCATATCTCTGGAATCCAGAGGATATGTT 120
|||||
DB 62629 TGATGCTTTTGGCTTGTATTCTATGGGACCATATCTCTGGAATCCAGAGGATATGTT 62570
|||||
QY 121 AGCTAGTCAAGATTGGGGGAAAATACCGGATGATGATCAGAGGAGGCTCCACTGTTTT 180
|||||
DB 62569 AGCTAGTCAAGATTGGGGGAAAATACCGGATGATGATCAGAGGAGGCTCCACTGTTTT 62510
|||||
QY 181 AAGCCGAAAGATGGCAAACTTTTGCAAGTTTTCATCTTTTCAG 224
|||||
DB 62509 AAGCCGAAAGATGGCAAACTTTTGCAAGTTTTCATCTTTTCAG 62466
|||||

RESULT 6
ATHS52638
LOCUS
Arabidopsis thaliana T-DNA flanking sequence, left border, clone
340812.
ACCESSION AJ552638
VERSION
left border; T-DNA flanking sequence.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1
REFERENCE
AUTHORS
Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny, A.
T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)
22363535
MEDLINE
12446565
PUBMED
2 (bases 1 to 695)
REFERENCE
AUTHORS
Balzerque, S.
JOURNAL
TITLE
Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT
PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).
FEATURES
Location/Qualifiers
1..695
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Wassiljewskij4"
/db_xref="taxon:3702"
/clone="340812"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
1..695
/note="T-DNA flanking sequence
left border"
ORIGIN
Query Match 35.9%; Score 180; DB 8; Length 695;
Best Local Similarity 100.0%; Pred. No. 7.7e-41;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTCATAACACTCCAGTCATCATGTAATCTTCTGGTACAGATGGTACATTGGATGGAAA 60
DB 516 AGTCATAACACTCCAGTCATCATGTAATCTTCTGGTACAGATGGTACATTGGATGGAAA 575
QY 61 TGATGCTTTTGGCTTGATTTATGGCGAGCCCATATCTCTGGAATCCAGAGGATATGTT 120
DB 576 TGATGCTTTTGGCTTGATTTATGGCGAGCCCATATCTCTGGAATCCAGAGGATATGTT 635
QY 121 AGCTAGTCAAGATTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTT 180
DB 636 AGCTAGTCAAGATTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTT 695
RESULT 7
LOCUS
AK068790 4027 bp mRNA linear PLN 24-JUL-2003
DEFINITION
Oryza sativa (japonica cultivar-group) cDNA clone:J013163J20, full insert sequence.
ACCESSION
AK068790
VERSION
AK068790.1 GI:32978815
KEYWORDS
FLI CDNA: CAP trapper.
SOURCE
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS
The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team:;

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariakawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, K., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
Science 301 (5631), 376-379 (2003)
22752273
PUBMED
2 (bases 1 to 4027)
REFERENCE
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Haehizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariakawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuka, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, X., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan [E-mail: skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T. and Yamamoto, M.
FAIS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariakawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,


```

Center clone name: 458N12
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 206439 bases at least Q40
Consensus quality: 206998 bases at least Q30
Consensus quality: 207487 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 208978; sum-of-contents
Quality coverage: 8.49x in Q20 bases; agarose-fp
Quality coverage: 7.19x in Q20 bases; sum-of-contents
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 21964: contig of 21964 bp in length
* 21965: gap of unknown length
* 22064: gap of unknown length
* 22065: contig of 31906 bp in length
* 53970: gap of unknown length
* 53971: gap of unknown length
* 54070: gap of unknown length
* 54071: contig of 35678 bp in length
* 89748: contig of 35678 bp in length
* 89749: gap of unknown length
* 89848: gap of unknown length
* 89849: contig of 49701 bp in length
* 139549: gap of unknown length
* 139550: contig of 49701 bp in length
* 139550: gap of unknown length
* 139550: contig of 49701 bp in length
* 139550: contig of 49701 bp in length.
*
FEATURES
source
1. .209378
/organism="Callithrix jacchus"
/mol_type="genomic DNA"
/db_xref="taxon:9483"
/clone_lib="CH259-458N12"
/notes="BAC resource: http://bacpac.chori.org/"
misc_feature
1. .21964
/notes="assembly_fragment"
22065. .53970
/notes="assembly_fragment"
54071. .89748
/notes="assembly_fragment"
vector_side:right
clone_end:T7
89849. .139549
/notes="assembly_fragment"
clone_end:SP6
vector_side:right
139550. .209378
/notes="assembly_fragment"
ORIGIN
Query Match 8.1%; Score 40.4; DB 2; Length 209378;
Best Local Similarity 59.6%; Pred. No. 1.5;
Matches 68; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 98 TCTGGATCCAGGATGTTAGCTAGTCAGATGGGGGAAATACCGGATGATCA 157
Db 107355 TCATTAAATCCATGCTATGATGCTGGTGAGATGGGAGGATATCAAGGAGACTTCA 107296
QY 158 CAGAGGAGGCTCCACACTGTTTAAAGCCGAGATGCGCAAAACTTTGCCAAGTT 211
Db 107295 CAGAGTAGTTGACACACAAATGAAATTTAAAGATGCAAAAGAGATTTCTAGAT 107242
RESULT 13
BX294103/c 193969 bp DNA linear VRT 04-JUL-2003
LOCUS Zebrafish DNA sequence from clone CH211-257A8 in linkage group 9,
DEFINITION complete sequence.
ACCESSION BX294103

```

```

VERSION BX294103.11 GI:32452291
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 193969)
Wood,J.
Direct Submission
Submitted (04-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jul 4, 2003 this sequence version replaced gi:32451361.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SW, SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/Clone-derived
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'dr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-257A8
is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
FEATURES
source
1. .193969
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-257A8"
/clone_lib="CHORI-211"
ORIGIN
Query Match 7.9%; Score 39.8; DB 5; Length 193969;
Best Local Similarity 49.3%; Pred. No. 2.2;
Matches 104; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 64 TGGTTTGGCTGTATTTCTATGGGAGCCATATCTCTGGAATCCAGAGGATGTGTAGC 123
Db 13010 TGGTTATGTTGGATTTCTTTTGTAGCTACTTTGGTGGCTCTTGAGATGTTGAAAG 12951
QY 124 TAGTCAAGATGGGGGAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTAA 183
Db 12950 TATCTACGATAGGATAGAGGATAGATGATGATGATGATGATGATGATGATGATG 12891
QY 184 GCGAAGATGGCAAACTTTGCCAAGTTTTCATCTTCAGATGCTTGCACAGCATGGT 243

```


Db 12890 AATAATCTGGAACCGATGAAAAAATATGGAAAGAAATTATGCAAGCAAAAGTTTTTT 12831
 QY 244 CGGAATTTTCTCGAATATGTTATTGAAAAAT 274
 Db 12830 TTGTAATTTTCTGTAATATTTCTCTGAGAGT 12800

RESULT 14

AC149661

LOCUS

DEFINITION

AC149661 154492 bp DNA linear HTG 04-SEP-2004
 Bos taurus clone CH240-65F15, WORKING DRAFT SEQUENCE, 22 unordered
 pieces.

ACCESSION

AC149661.2

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

Bos taurus (cow)

SOURCE

Bos taurus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

1 (bases 1 to 154492)

REFERENCE

AUTHORS

Muzny D., Metzker M., Adams C., Agbai II O., Allen C.,
 Albrooks S., Archer P., Arredondo H., Bandaranaike D., Bangura L.,
 Beltran B., Beltran R., Bereducci A., Biswal K., Blyth P.,
 Bonham H., Buhay C., Burch P., Cadoree I., Canada A., Cardenas V.,
 Carter K., Cavazos I., Chacko J., Chahrour M., Chavez D., Chen A.,
 Chen G., Chen R., Cheng M.-T., Chu J., Clerc K., Cockrell R.,
 Coyle M., Cree A., Curry S., Dai W., Davila M.L., Davis C.,
 Davy-Carroll L., De Anda C., Delgado O., Denson S., Deramo C.,
 Ding Y., Dinh H., Donlin J., McCauley S., Dugan-Kocha S., Dunn A.,
 Durbin K., Dziuda D., Egan A., Escotto M., Espinosa V., Eugene C.,
 Fa M., Fernandez S., Fernando P., Flagg N., Forbes L., Foster P.,
 Fowler G., Fu Q., Fuh E., Garcia A., Garcia R., Garner T.,
 Gaslin C., Gench S., Ghose S., Gill R., Gonzalez D.,
 Gonzalez-Garay M., Guevara W., Holder M., Haaland W., Haeblerlein K.,
 Hall B., Hamid H., Hamilton K., Harbes B., Harris R., Havlak P.,
 Hawes A., Hawkins E., Hayes S., Hemphill L., Hernandez J.,
 Hines S., Hitchens M., Hodgson A., Hogue M., Hollins B.,
 Howell L.T., Hulyk S., Hume J., Imo K., Jackson A., Jackson L.,
 Jacob L., Jiang H., Johnson B., Johnson R., Kalafus K., Kelly S.,
 Keys T., Khan Z., King L., Kovar C., Kowis A., Kowis C., Lara F.,
 Leal S., Lee K., Lee S., Legall F.I., Lemon S., Lewis L., Li B.,
 Li Y., Li Z., Linnell M., Liu W., Liu Y.-S., Liu Y., Liyanage D.,
 London P., Lopez J., Lorenshewal L., Lozado R., Luk T., Madu R.,
 Maheshwari M., Mahoney C., Mallory K., Mansouri D., Martinez E.,
 McClelland H., McPherson J., Mercadao C., Milosavljevic A.,
 Minja E., Morgan M., Morris S., Munidasa M., Murray D.,
 Nazareth L., Ngo D., Nguyen N., Norwig-Eastaugh E., Nott A.,
 Nwaokemele O., Obregon M., Ochi-Okorie C., Odeh E., Okwuonu G.,
 Okwuonu K., Parker D., Pasternak S., Patel B., Patel V., Paul H.,
 Perez A., Perez L., Petrosino J., Pham T., Primus E., Pu L.-L.,
 Puazo M., Qin X., Quinn A., Quiroz J., Rabata D., Rachlin E.,
 Reigh R., Ren Y., Reuter M., Richards S., Rives C., Rodriguez F.,
 Rojas A., Ruiz S.J., Sana M., Sanders W., Santibanez J., Santos R.,
 Savary G., Scherer S., Shen H., Shen Y., Sisson I., Sneed A.,
 Sodergren E., Song X.-Z., Sorelle R., Svatek A., Taylor E.,
 Taylor T., Thomas N., Thorn R., Thornton R., Trejos Z., Usmani K.,
 Vazco C., Verdusco D., Villaseana D., Virk D., Volkov A.,
 Waldron L., Walker B., Wang Q., Wang S., Warren J., Wei X.,
 Wheeler D., Williams G., Williams R., Worley K., Wright R., Wu J.,
 Yakub S., Yan K., Yuan Y., Yu F., Zhang J., Zhang L., Zhang Z.,
 Zhou J., Weinstock G. and Gibbs R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Submitted (19-JUN-2004) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

JOURNAL

REFERENCE

AUTHORS

TITLE

3 (bases 1 to 154492)

Direct Submission

JOURNAL

COMMENT

Submitted (04-SEP-2004) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 23, 2004 this sequence version replaced gi:48958564.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help.tmc.edu

----- Project Information
 Center project name: FDLQ
 Center clone name: CH240-65F15

----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 154691 bases at least Q40
 Consensus quality: 157632 bases at least Q30
 Consensus quality: 160679 bases at least Q20

Estimated insert size: 162987; sum-of-contigs estimation
 Estimated insert size: 162510; agarose-fp estimation
 Quality coverage: 3x in Q20 bases; agarose-fp estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

3549: contig of 3549 bp in length
 3649: gap of unknown length
 7528: contig of 3879 bp in length
 7628: gap of unknown length
 14764: contig of 7136 bp in length
 14964: gap of unknown length
 14765: 19106: contig of 4242 bp in length
 14865: 19106: gap of unknown length
 19106: gap of unknown length
 19206: gap of unknown length
 28997: contig of 9791 bp in length
 28997: gap of unknown length
 34047: contig of 4950 bp in length
 34147: gap of unknown length
 40354: contig of 6207 bp in length
 40454: gap of unknown length
 40355: 45444: contig of 5290 bp in length
 45444: gap of unknown length
 45745: 55107: contig of 9263 bp in length
 45845: 55207: gap of unknown length
 55108: 58904: contig of 3597 bp in length
 58904: contig of 3597 bp in length
 58905: 61881: gap of unknown length
 61881: gap of unknown length
 61882: 68294: contig of 6713 bp in length
 68294: gap of unknown length
 68395: 77182: contig of 8788 bp in length
 77182: gap of unknown length
 77283: 82073: contig of 4791 bp in length
 82073: gap of unknown length
 82074: 88358: contig of 6185 bp in length
 88358: gap of unknown length
 88459: 98110: contig of 9652 bp in length
 98110: gap of unknown length
 98111: 100840: contig of 2630 bp in length
 100840: gap of unknown length
 100841: 100940: contig of 6090 bp in length
 100941: 107130: gap of unknown length
 107131: 138133: contig of 31003 bp in length
 138133: gap of unknown length
 138134: 144358: contig of 6125 bp in length
 144358: contig of 6125 bp in length

* 144359 144458: gap of unknown length
* 144459 151604: contig of 7146 bp in length
* 151605 151704: gap of unknown length
* 151705 154492: contig of 2788 bp in length.

FEATURES
 source
 1..154492
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="CH240-65F15"

ORIGIN
Query Match 7.9%; Score 39.4; DB 2; Length 154492;
Best Local Similarity 50.3%; Pred. No. 2.9;
Matches 97; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 85 GGGCAGCCATATCTCTGGAATCCAGAGGATATCTAGCTAGTCAAGATTGGGGGAAAT 144
 |||||
Db 35083 GGGAGGCTGTGTGTGCTGCAGTCCATGGAATGCAGAGTCGACATGACTGAGTGACTGA 35142
 |||||

Qy 145 ACCGGATGAATCACAGAGGAGGCTCCACACTGTTTTTAAAGCCGAGATGGCAAACTTTG 204
 |||||
Db 35143 ACTGAACCTGAACCTGAACCTGAGTGTCTGACTGAATTAACCTGGGACACCACCTGTTG 35202
 |||||

Qy 205 CCAAGTTTTCATCTTCAGATGCTTCGACAGCATGCTCGAAATTTTCTCGAATATGT 264
 |||||
Db 35203 CAGAGACTTGTGTTGTATGTGAACACTTACATGATGCTTCAAAAATATTTTGACTGTGT 35262
 |||||

Qy 265 TATTGAATACAC 277
 |||||
Db 35263 GAGTAAAGTTAAC 35275
 |||||

RESULT 15
AE013102/c

LOCUS 10303 bp DNA linear BCT 10-JUN-2004
DEFINITION Thermoanaerobacter tengcongensis MB4, section 129 of 244 of the
 complete genome.
ACCESSION AE013102 AE008691
VERSION AE013102.1 GI:20516444
KEYWORDS
SOURCE Thermoanaerobacter tengcongensis MB4
 Thermoanaerobacter tengcongensis MB4
 Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 Thermoanaerobacteriaceae; Thermoanaerobacter.
REFERENCE
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J.,
 Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L.,
 Tan, H., Chen, R., Wang, J., Yu, J., and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
MEDLINE Genome Res. 12 (5), 689-700 (2002)
PUBMED 21992816
 11997336

REFERENCE
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and
 Bioinformatics Center, Institute of Genetics and Development,
 Chinese Academy of Sciences, Beijing Airport Industrial Zone B6,
 Beijing 101300, China
 3 (bases 1 to 10303)
 Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of
 Sciences, Beijing 100101, China
 4 (bases 1 to 10303)
 Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy
 of Sciences, Beijing 100080, China
 1..10303
 /organism="Thermoanaerobacter tengcongensis MB4"

FEATURES
 source
 1..10303

/mol_type="genomic DNA"
/strain="MB4"
/db_xref="taxon:273068"
complement(24..1284)
/gene="FlgE3"
/locus_tag="TTE1433"
complement(24..1271)
/gene="FlgE3"
/locus_tag="TTE1433"
/notes="Best Blastp hit = gi|2494528|sp|Q44767|FLGE_BORBU
FLAGELLAR HOOK PROTEIN FLGE gi|7463172|pir|C70135_
FLAGELLAR HOOK PROTEIN (flgE) homolog - Lyme disease
spirochete gi|1165267|gb|AA85606.1| (U43739) FlgE
[Borrelia burgdorferi] gi|2688183|gb|AAC6665.1|
[AE001137] flagellar hook protein (flgE) [Borrelia
burgdorferi], score 278, E-value 1.00E-73"
/codon_start=1
/transl_table=11
/product="Flagellar basal body and hook proteins"
/protein_id="AA024655.1"
/db_xref="GI:20516445"
/translations="MLRSMYSAISGLQAHQARLDVIGNIANVNTVGYKASMTFKEI
FSOTIKGASAPQNGGGTNPQOIGLGVAIASIDTLFTRGGAORTDNPDTLSIDNGGPF
IVNGSILYTRAGNFSFNSGLDVTDPGKVLGWMSTDKTNTVDTGNLVPISLKNW
FGMEPTTQLEIGGNLNASTIGQSISINIIVIDTQGGSHVATITFROQDPTNWS
VSSQDPYISVSVAGSTAITFGADGKIAPNTQATGLTFNNMTAVTNAQIGAPVNDIS
KLMTFTETDLRELSKNGNEAGSLQISINIDKYGVGSIYNGRRQVIGQIALADFPNP
MGLKVGNTWFIINTVNSGEPMIGAAGSGTGRSINPGTLEMSNVNLANFEVDMITTQRG
QANAKVITVSDMLQDLVNMKE"
complement(1167..1259)
/gene="FlgE3"
/locus_tag="TTE1433"
/notes="Pfam match to entry flg_bb rod, Flagella basal body
rod protein, score 58.3, E-value 1.30E-13"
complement(1280..1284)
/gene="FlgE3"
/locus_tag="TTE1433"
complement(1352..1712)
/locus_tag="TTE1434"
complement(1352..1702)
/locus_tag="TTE1434"
/notes="Best Blastp hit = gi|10175070|dbj|BAB06169.1|
[AP001515] BH2450
unknown [Bacillus halodurans], score 75.5, E-value
2.00E-13"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA024656.1"
/db_xref="GI:20516446"
/translations="MKVHFNSLNGPSIQKPPQPSYGBAFKQVFEKLAIEIKFSKSLR
MEARNVKITPDEYKLLKLEAVEKASSKGKINDLIIMDNKAPVNLKSRVTITAMDESML
KNSVFTNIDGNVIL"
complement(1708..1712)
/locus_tag="TTE1434"
complement(1712..2122)
/gene="FlgD"
/locus_tag="TTE1435"
complement(1712..2107)
/gene="FlgD"
/locus_tag="TTE1435"
/notes="Best Blastp hit = gi|7459749|pir|D70135 flagellar
hook assembly protein (flgD) homolog - Lyme disease
spirochete gi|1165268|gb|AA85607.1| (U43739) YlxG
[Borrelia burgdorferi] gi|1196325|gb|AB851417.1| (U76303)
flagellar hook assembly protein [Borrelia burgdorferi]
gi|2688182|gb|AAC6664.1| (AE001137) flagellar hook
assembly protein (flgD) [Borrelia burgdorferi], score
85.1, E-value 3.00E-16"
/codon_start=1
/transl_table=11
/product="Flagellar hook capping protein"
/protein_id="AA024657.1"

```
/db_xref="GI:20516447"
/translation="MNVTNYSVLPPSNQNKVYTKSKLGDPLMLLVTLQLRNQDPLN
PYDKFEFLAQFSALEQMNLSFNSVKAIALIGKNVYATINDNGNGYKEVFGKV
DVAIRENGQYFLKVKGLDPLDRAVATE"
complement (2118..2122)
/gene="FlgD"
/locus_tag="TTE1435"
complement (2125..3349)
/locus_tag="TTE1436"
complement (2125..3336)
/locus_tag="TTE1436"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM24658.1"
/db_xref="GI:20516448"
/translation="MSEQVVQVELLPFLKAVDSKMAKGKNGKVSFLDLLSELTE
TREGNKENPLNKKKEEPLSQKNGLKNVEIPQEKYKEDKDIMEDLNFPALMQITIL
SLETISSEGASDFEKVRKLEVALQGFIERNFYFKELAKKISDFLKNFNIEISPE
VIERHKAKVDLDLPFLQDLNQDFAENVQKNDKTSQDKIDKBAFIAAKEABKE
KTEKSFQKQFVFPEKNGKPVSNLTYSIKKNDPVDRLFRQIVDVFVAKEKAS
SVTNLKPILGLQISLKSIDGNIIVATIVTSEKTHQIESNLSLQALDLKGIKI
ESVNVADNKFOTPTQYTGEGNVQSGFYGENENLRYFHYGSKYLPSELDLTSQNY
YFKDGHDLMA"
complement (3345..3349)
/locus_tag="TTE1436"
complement (3352..3944)
/locus_tag="TTE1437"
complement (3352..3930)
/locus_tag="TTE1437"
/notes="Best Blastp hit = gi|10175073|dbj|BAB06172.1|
(AP001515) BH2453
unknown conserved protein in B. subtilis [Bacillus
halodurans], score 62.8, E-value 3.00E-09"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM24659.1"
/db_xref="GI:20516449"
/translation="MONEQVQEKRSKAKWIIALILVILSSVLVYFNIGSOVFL
TKQISKIPLLKGVSPKPTKEDAVIAKNEELKKEATISNYEEKIKQNAQIELK
KQIELQNEKLTLKQFESRQNSLKDIAITYQNMDPQNAQAQILNMSDEVIKILTSM
NKDSAKIMEALKPEKAAKITNMLKKSATWVP"
complement (3940..3944)
/locus_tag="TTE1437"
complement (3943..4397)
/locus_tag="TTE1438"
complement (3943..4386)
/locus_tag="TTE1438"
/notes="Best Blastp hit = gi|7462657|pir||F72368
hypothetical protein - Thermotoga maritima (strain MSB8)
gi|4981024|gb|AAD35597.1|AE001727.9 (AE001727)
hypothetical protein [Thermotoga maritima], score 57,
E-value 1.00E-07"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM24660.1"
/db_xref="GI:20516450"
/translation="MKKFEFTLOSVLNKEQSEKIEKENLAKIMKEIEREREKENLK
KHLQEVTKRAKEVEEGTLMYKLAETIYIMKIREMKQANTILKLEAKEAKIREGL
LKVSKEKALENLKQRFSEYLLNLEQSRVIDEHVSRYKVAKS"
complement (4392..5716)
/gene="Flil"
/locus_tag="TTE1439"
complement (4392..5705)
/gene="Flil"
/locus_tag="TTE1439"
/notes="Best Blastp hit = gi|10175075|dbj|BAB06174.1|
(AP001515) flagellar-specific ATP synthase [Bacillus
halodurans], score 529, E-value 1.00E-149"
/codon_start=1
/transl_table=11
```

```
/product="Flagellar biosynthesis/type III secretory
pathway ATPase"
/protein_id="AAM24661.1"
/db_xref="GI:20516451"
/translation="MSSGVLEKYSKILREKRLVGYGVKVSQVIGLTIESVGPLSNIGE
ICVYKIDGNEVLAERVGVKKEKYLMPLGNMEGIGPGSKVIATGOTLKVNVKSLLG
RVLDGLGNPIDGKPLKYEKSI PVNNTPPDPLERKRIRVEMPLGKAIKDGLATCGKQ
RIGIFAGSGVGKSTLLGMARNKADLNIALIGRGRVNEFIEKDLGEEGLKRSVV
VWATSDTPALVRVKGMATATAEYFRDQGLDVLMDISITRFAMAQRIGUSIGEAP
VSRGTPSVFVLPKLLERSGSSQKSI TALYTVLVGDLDLNEPTADAVRGILDHIV
LSRKLANKNHYPAIDVLASVSRVINDIITDEHKELVAKPKNLTATVAEADLINIGAY
NFGTNPKIDAEILELNEKNSEFLRQKIDEAYDEFTTKLLAKSLER"
complement (4393..4397)
/locus_tag="TTE1438"
complement (4623..5432)
/gene="Flil"

RBS
misc_feature

Query Match 7.8%; Score 39.2; DB 1; Length 10303;
Best Local Similarity 47.9%; Pred. No. 3.4; Mismatches 123; Indels 0; Gaps 0;
Matches 113; Conservative 0;

QY 207 AAGTTTTCATCTTTTCAGATGCTTGCAAGCATGCTCGCAAGATTTTCTCGAATATGTTA 266
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3158 AAGAATTCCTCTCAAAAAATGGCTAAAAAATGTAGAAATTTTCAAGAAAAAGTAA 3099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 267 TTGAAAAATCAGCAATCTTACGAAGAGCCAGCCACTACTTTTCAGGATTCAGATAGCCC 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3098 AAGAAGATAAGGATATAATGGAGGATTTAAACAAATTTTATCCAGCATTAATGCAGACTA 3039
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 327 TCAGTTGGATTGCAGCCTTGTGTAAGCAAAATCTTAGCCACAAAGAAATCTCTCGTCC 386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3038 TAACTTTGAGTTTAGAGCAGATTTCAAGTGAAAAAGGTCCATCAGACTTTTGAAGAGTCA 2979
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 387 GTGCAAAATTCGAATTAGCTTTCAAAATGCTCTAGAGTAGAGGTGGATTATATTTAT 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2978 GAGAAAAATTAGAGTGGCTCTACAAGGTTTTTATAAAGAAAGGAATTTTACTTTT 2923
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 12, 2005, 05:59:46
Job time : 2541.52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 382.755 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929c-2_COPY_3000_3500

Perfect score: 501

Sequence: 1 agcataactccatgcattcat.....acacagggttgagttcgca 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	6571	4	AAA89354 Arabidops
2	224	44.7	10329	4	AAA89385 Arabidops
3	224	44.7	10329	4	AAA89353 Arabidops
C 4	63.4	12.7	417	4	AAA89384 Brassica
5	57.6	11.5	515	4	AAA89381 Brassica
C 6	42.6	8.5	2000	8	ADA71938 Rice gene
7	37.4	7.5	2000	8	ADA71938 Rice gene
C 8	36.8	7.3	21441	4	AAI62667 Human bre
C 9	36.8	7.3	21441	4	AAI62667 Human bre
C 10	35.8	7.1	458	9	ACH24966 Human adu
11	35.6	7.1	110000	12	ADQ97960 Human can
12	35.4	7.1	3160	10	ADF69148 Human can
C 13	34.8	6.9	534	13	ACN55834 Cotton an
C 14	34.6	6.9	110000	2	AAV21209_03 Continuation (4 of
15	34	6.8	10113	4	AAS32838 Human gen
16	34	6.8	11655	4	AAS32839 Human gen
17	33.6	6.7	88493	13	ABD33522 Murine ca
18	33.2	6.6	1449	8	ACA21089 Prokaryot
C 19	33.2	6.6	5879	6	ABL32269 Human imm
C 20	33	6.6	1224	8	ACN45078 Prokaryot

21	33	6.6	110000	6	ABA03041_00
C 22	32.8	6.5	67832	9	ADA02801_00
C 23	32.8	6.5	67832	10	ADB72539 Mouse itk
C 24	32.8	6.5	67832	10	ADC85281 Mouse itk
C 25	32.8	6.5	67832	12	ADM74396 Murine ca
26	32.8	6.5	135827	13	ABD33219 Human can
27	32.6	6.5	469	6	ABV96113 Human pan
28	32.6	6.5	744	10	ADF72223 Human ZLM
29	32.6	6.5	2213	4	ABL21260 Drosophil
30	32.6	6.5	2623	9	ABT43708 Molecule
31	32.6	6.5	2719	11	ADM02508 Human cdn
32	32.6	6.5	2750	3	AAC69110 Human sec
33	32.6	6.5	2752	3	AAC69119 Human sec
C 34	32.6	6.5	58320	13	ABD33125 Human can
C 35	32.4	6.5	400	8	ABX91579 Murine ge
36	32.4	6.5	3107	13	ADS89386 Oligonuc1
37	32.4	6.5	3107	13	ADS89660 Oligonuc1
38	32.4	6.5	3501	13	ADS89484 Oligonuc1
39	32.4	6.5	3501	13	ADS89758 Oligonuc1
40	32.4	6.5	6107	6	ABK31431 Signal tr
41	32.4	6.5	6107	6	ABL70390 Chemical
42	32.4	6.5	6107	6	AAS61342 Human gen
C 43	32.4	6.5	349901	10	ADC86940 Human GPC
C 44	32.4	6.5	349938	10	ADC87621 Human GPC
C 45	32.2	6.4	220224	11	ACN44702 Human gen

ALIGNMENTS

RESULT 1
AAA89354
ID AAA89354 standard; cdna; 6571 BP.
XX
AC AAA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 310..6315
FT /*tag= a
XX
FN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PP 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
(NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX
PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 6; Page 22-32; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, in particular

CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression
XX
SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 4; Length 6571;
Best Local Similarity 100.0%; Pred. No. 3.8e-143;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTCATAACCTCCAGTCATCATGTAATCTCTGGTACAGATGGTACATTTGGATGAAA 60
Db 3000 AGTCATAACCTCCAGTCATCATGTAATCTCTGGTACAGATGGTACATTTGGATGAAA 3059
Qy 61 TGATGCTTTTGGCTTGATTTCTATGGCAGCCATCTCTGGAAATCCAGAGGATATGTT 120
Db 3060 TGATGCTTTTGGCTTGATTTCTATGGCAGCCATCTCTGGAAATCCAGAGGATATGTT 3119
Qy 121 AGCTAGTCAAGATTGGGGGAAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT 180
Db 3120 AGCTAGTCAAGATTGGGGGAAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT 3179
Qy 181 AAGCCGAAGATGGCAAAACTTTGCCAAGTTTTCATCTTTTCAGATGCTTGCACAAGCAT 240
Db 3180 AAGCCGAAGATGGCAAAACTTTGCCAAGTTTTCATCTTTTCAGATGCTTGCACAAGCAT 3239
Qy 241 GGTCCGAAATTTCTCGAATATGTTATTGAAATCACCAGATCTTACAGATGCTTGCACAAGCAT 300
Db 3240 GGTCCGAAATTTCTCGAATATGTTATTGAAATCACCAGATCTTACAGATGCTTGCACAAGCAT 3299
Qy 301 TACTTTTCAGGCATTCAGATAGCCCTTCAGTTGGATTGCGCCCTTGTGTAAGCAAAAT 360
Db 3300 TACTTTTCAGGCATTCAGATAGCCCTTCAGTTGGATTGCGCCCTTGTGTAAGCAAAAT 3359
Qy 361 TCTTAGCCACAAGAAATCTCTGGTCCGTGCAAAATTCGAATTCAGATGCTTGCACAAGCAT 420
Db 3360 TCTTAGCCACAAGAAATCTCTGGTCCGTGCAAAATTCGAATTCAGATGCTTGCACAAGCAT 3419
Qy 421 AGTAGAGGTGATTTATTTTATTCGATATGTTCTCTGCATGAAGAGTCTGTTCTCGGAGCA 480
Db 3420 AGTAGAGGTGATTTATTTTATTCGATATGTTCTCTGCATGAAGAGTCTGTTCTCGGAGCA 3479
Qy 481 TACACAGGTTTGCAGTTTCA 501
Db 3480 TACACAGGTTTGCAGTTTCA 3500

RESULT 2
AAA89385
ID AAA89385 standard; DNA; 10329 BP.
XX
AC AAA89385;
XX
DT 04-NOV-2004 (revised)
XX 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Arabidopsis thaliana.
XX

Key Location/Qualifiers
exon 1..1008
FT /tag= a
FT /number= 1
FT 1009..1295
FT /tag= b
FT /number= 1
FT 1296..2550
FT /tag= c
FT /number= 2
FT 1301..9808
FT /tag= d
FT /product= "Columbia protein"
FT /note= "This sequence contains 15 introns."
FT 2551..2673
FT /tag= e
FT /number= 2
FT 2674..2752
FT /tag= f
FT /number= 3
FT 2753..2867
FT /tag= g
FT /number= 3
FT 2868..3113
FT /tag= h
FT /number= 4
FT 3114..3506
FT /tag= i
FT /number= 4
FT 3507..3680
FT /tag= j
FT /number= 5
FT 3681..3973
FT /tag= k
FT /number= 5
FT 3974..4895
FT /tag= l
FT /number= 6
FT 4338
FT /tag= m
FT /note= "ecotype Zurich"
FT 4896..4975
FT /tag= n
FT /number= 6
FT 4976..5217
FT /tag= o
FT /number= 7
FT 5218..5777
FT /tag= p
FT /number= 7
FT 5778..5882
FT /tag= q
FT /number= 8
FT 5883..6082
FT /tag= r
FT /number= 8
FT 6083..7480
FT /tag= s
FT /number= 9
FT 6721
FT /tag= t
FT /note= "ecotype Zurich"
FT 7481..7615
FT /tag= u
FT /number= 9
FT 7616..7771
FT /tag= v
FT /number= 10
FT 7772..7914
FT /tag= w
FT /number= 10
FT 7915..8070
FT /tag= x
FT

FT intron /number= 11
 FT 8071..8153
 FT /*tag= y
 FT /number= 11
 FT 8154..8318
 FT /*tag= z
 FT /number= 12
 FT 8319..8451
 FT /*tag= aa
 FT /number= 12
 FT 8452..8629
 FT /*tag= ab
 FT /number= 13
 FT 8630..8718
 FT /*tag= ac
 FT /number= 13
 FT 8719..8918
 FT /*tag= ad
 FT /number= 14
 FT 8919..9000
 FT /*tag= ae
 FT /number= 14
 FT 9001..9211
 FT /*tag= af
 FT /number= 15
 FT 9212..9284
 FT /*tag= ag
 FT /number= 15
 FT 9285..10329
 FT /*tag= ah
 FT /number= 16

XX WO200100801-A2.

PN 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.

XX 23-JUN-1999; 99GB-00014623.

XX (NOVS) NOVARTIS AG.

XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;

XX WPI; 2001-137952/14.

XX P-PSDB; AAB20063.

XX Novel gene encoding a protein that controls gene silencing, in particular silencing of plant genes.

XX Claim 5; Page; 48pp; English.

XX The present sequence is that of an Arabidopsis thaliana ecotype Columbia gene (MOM) involved in epigenetic gene silencing. It encodes a 2001-amino acid protein with similarity with ATPase/helicase proteins of the SWI2/SNF2 family, and which controls gene silencing, particularly plant gene silencing. Trans-acting modified loci were identified by T-DNA insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which carries a heritable inactivated, methylated hygromycin resistance gene (hyg). Mutation of a silencing modified locus results in release of silencing of the hyg gene and restores hygromycin resistance. Plants homozygous for the silent resistance gene were subjected to transformation with a bar selectable marker gene under control of the T-DNA 1'-2' dual promoter. Transformants were selected and their progeny screened for hygromycin resistance. The mutant phenotype (hygromycin resistance) was screened for genetic co-segregation with a specific T-DNA insert. Cloning of the tagged gene allowed characterization of the present sequence. 2 Allelic differences were found in the corresponding gene of A. thaliana ecotype Zurich, occurring at nucleotide positions 4338 (T instead of A) and 6721 (G instead of T). Gene silencing is useful as a molecular tool for regulating gene expression. Note: the present sequence is not shown in the specification but is derived from the A.

CC thaliana mutant silencer gene sequence given in AAA89353
 CC Revised record issued on 04-NOV-2004 : Correction to feature table key
 CC XX Sequence 10329 BP; 3106 A; 1934 C; 2141 G; 3148 T; 0 U; 0 Other;
 SQ Query Match 44.7%; Score 224; DB 4; Length 10329;
 Best Local Similarity 100.0%; Pred. No. 7.5e-58;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 AGTCATAACACTCCAGTCATCATGTAATCTTCGTACAGATGATGATGATGATAA 60
 Db 4995 AGTCATAACACTCCAGTCATCATGTAATCTTCGTACAGATGATGATGATAA 5054
 Qy 61 TGATGCTTTTGGCTTGATTTCTATGGCAGCCATATCTCTGGAATCCAGAGGATATGTT 120
 Db 5055 TGATGCTTTTGGCTTGATTTCTATGGCAGCCATATCTCTGGAATCCAGAGGATATGTT 5114
 Qy 121 AGCTAGTCAGATTTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT 180
 Db 5115 AGCTAGTCAGATTTGGGGGAAATACCGGATGAATCACAGAGGAGGCTCCACACTGTTTT 5174
 Qy 181 AAAGCCGAAGATGCAAAACTTTGCCAAGTTTTTGCATCTTTTCAG 224
 Db 5175 AAAGCCGAAGATGCAAAACTTTGCCAAGTTTTTGCATCTTTTCAG 5218
 RESULT 3
 AAA89353
 ID AAA89353 standard; DNA; 10329 BP.
 XX AC AAA89353;
 DT 04-NOV-2004 (revised)
 DT 23-APR-2001 (first entry)
 XX Arabidopsis thaliana gene involved in epigenetic gene silencing.
 XX Gene silencing; silencing gene; MOM; ds.
 OS Arabidopsis thaliana.
 FH Key Location/Qualifiers
 FT exon 1..1008
 FT /*tag= a
 FT /number= 1
 FT intron 1009..1295
 FT /*tag= b
 FT /number= 1
 FT exon 1296..2550
 FT /*tag= c
 FT /number= 2
 FT CDS 1301..9808
 FT /*tag= d
 FT /product= "MOM protein"
 FT /note= "This sequence contains 15 introns."
 FT intron 2551..2673
 FT /*tag= e
 FT /number= 2
 FT exon 2674..2752
 FT /*tag= f
 FT /number= 3
 FT intron 2753..2867
 FT /*tag= g
 FT /number= 3
 FT exon 2868..3113
 FT /*tag= h
 FT /number= 4
 FT intron 3114..3506
 FT /*tag= i
 FT /number= 4
 FT exon 3507..3680
 FT /*tag= j

FT intron /number= 5
 FT 3681.. .3973
 FT /*tag= k
 FT /number= 5
 FT 3974.. .4895
 FT /*tag= l
 FT /number= 6
 FT 4338
 FT /*tag= m
 FT /note= "ecotype Columbia"
 FT 4896.. .4975
 FT /*tag= n
 FT /number= 6
 FT 4976.. .5217
 FT /*tag= o
 FT /number= 7
 FT 5218.. .5777
 FT /*tag= p
 FT /number= 7
 FT 5778.. .5882
 FT /*tag= q
 FT /number= 8
 FT 5883.. .6082
 FT /*tag= r
 FT /number= 8
 FT 6083.. .7480
 FT /*tag= s
 FT /number= 9
 FT 6721
 FT /*tag= t
 FT /note= "ecotype Columbia"
 FT 7481.. .7615
 FT /*tag= u
 FT /number= 9
 FT 7616.. .7771
 FT /*tag= v
 FT /number= 10
 FT 7772.. .7914
 FT /*tag= w
 FT /number= 10
 FT 7915.. .8070
 FT /*tag= x
 FT /number= 11
 FT 8071.. .8153
 FT /*tag= y
 FT /number= 11
 FT 8154.. .8318
 FT /*tag= z
 FT /number= 12
 FT 8319.. .8451
 FT /*tag= aa
 FT /number= 12
 FT 8452.. .8629
 FT /*tag= ab
 FT /number= 13
 FT 8630.. .8718
 FT /*tag= ac
 FT /number= 13
 FT 8719.. .8918
 FT /*tag= ad
 FT /number= 14
 FT 8919.. .9000
 FT /*tag= ae
 FT /number= 14
 FT 9001.. .9211
 FT /*tag= af
 FT /number= 15
 FT 9212.. .9284
 FT /*tag= ag
 FT /number= 15
 FT 9285.. .10329
 FT /*tag= ah
 FT /number= 16

XX WO200100801-A2.
 XX 04-JAN-2001.
 PD
 XX 21-JUN-2000; 2000WO-EP005761.
 PF
 XX 23-JUN-1999; 99GB-00014623.
 PR
 XX (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 PA
 XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 PI
 XX WPI; 2001-137952/14.
 XX P-PSDB; AAB20062.
 DR
 XX Novel gene encoding a protein that controls gene silencing, in particular
 FT silencing of plant genes.
 FT
 XX Claim 6; Page 19-22; 48pp; English.
 PS
 XX The present sequence is that of an Arabidopsis thaliana gene (MOM)
 CC involved in epigenetic gene silencing. It encodes a 2001-amino acid
 CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
 CC the SWI2/SNF2 family, which controls gene silencing, particularly plant
 CC gene silencing. Trans-acting modified loci were identified by T-DNA
 CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
 CC carries a heritable inactivated, methylated hygromycin resistance gene
 CC (hyg). Mutation of a silencing modified locus results in release of
 CC silencing of the hyg gene and restores hygromycin resistance. Plants
 CC homozygous for the silent resistance gene were subjected to
 CC transformation with a bar selectable marker gene under control of the T-
 CC DNA 1'-2' dual promoter. Transformants were selected and their progeny
 CC screened for hygromycin resistance. The mutant phenotype (hygromycin
 CC resistance) was screened for genetic co-segregation with a specific T-DNA
 CC insert. Cloning of the tagged gene allowed characterization of the
 CC present sequence. 2 Allelic differences were found in the corresponding
 CC gene of A. thaliana ecotype Columbia, occurring at nucleotide positions
 CC 438 (A instead of T) and 6721 (T instead of G). Gene silencing is useful
 CC as a molecular tool for regulating gene expression
 CC
 CC Revised record issued on 04-NOV-2004 : Correction to feature table key
 XX
 SQ Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;
 Query Match 44.7%; Score 224; DB 4; Length 10329;
 Best Local Similarity 100.0%; Pred. No. 7.5e-58;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTCATAACACTCCAGTCATCATGTAATCTTCTGGTACAGATGGTACATTGGATGAAA 60
 Db |||||
 QY 4995 AGTCATAACACTCCAGTCATCATGTAATCTTCTGGTACAGATGGTACATTGGATGAAA 5054
 Db |||||
 QY 61 TGATGCTTTTGGCTTGTATTCTATGGGACCCATATCTCTGGAATCCACAGGATATGTT 120
 Db |||||
 QY 5055 TGATGCTTTTGGCTTGTATTCTATGGGACCCATATCTCTGGAATCCACAGGATATGTT 5114
 Db |||||
 QY 121 AGCTAGTCAAGATGGGGGAAAATACCGGATGAATCACAGAGGAGCTCCACACTGTTT 180
 Db |||||
 QY 5115 AGCTAGTCAAGATGGGGGAAAATACCGGATGAATCACAGAGGAGCTCCACACTGTTT 5174
 Db |||||
 QY 181 AAAGCCGAAGATGGCAAACTTTTGGCAAGTTTTCATCTTTTCAG 224
 Db |||||
 QY 5175 AAAGCCGAAGATGGCAAACTTTTGGCAAGTTTTCATCTTTTCAG 5218
 Db |||||
 RESULT 4
 ID AAA89384/C
 XX AAA89384 standard; DNA; 417 BP.
 XX AAA89384;
 XX

KW	Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;	PR	25-SEP-2000; 2000US-0234997P.
KW	ds...	PR	25-SEP-2000; 2000US-0234998P.
XX		PR	26-SEP-2000; 2000US-0235484P.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834P.
XX		PR	27-SEP-2000; 2000US-0235836P.
PN	WO200155324-A2.	PR	29-SEP-2000; 2000US-0236327P.
XX		PR	29-SEP-2000; 2000US-0236367P.
PD	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236368P.
XX		PR	29-SEP-2000; 2000US-0236369P.
PF		PR	29-SEP-2000; 2000US-0236370P.
XX		PR	02-OCT-2000; 2000US-0236802P.
XX	17-JAN-2001; 2001WO-US001344.	PR	02-OCT-2000; 2000US-0237037P.
PR		PR	02-OCT-2000; 2000US-0237038P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237039P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237040P.
PR	24-FEB-2000; 2000US-0184664P.	PR	13-OCT-2000; 2000US-0239935P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239937P.
PR	16-MAR-2000; 2000US-0189874P.	PR	20-OCT-2000; 2000US-0240960P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0241221P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241826P.
PR	19-MAY-2000; 2000US-0205551P.	PR	01-NOV-2000; 2000US-0244617P.
PR	07-JUN-2000; 2000US-0209467P.	PR	08-NOV-2000; 2000US-0246474P.
PR	28-JUN-2000; 2000US-0214886P.	PR	08-NOV-2000; 2000US-0246475P.
PR	30-JUN-2000; 2000US-0215135P.	PR	08-NOV-2000; 2000US-0246477P.
PR	07-JUL-2000; 2000US-0216647P.	PR	08-NOV-2000; 2000US-0246478P.
PR	07-JUL-2000; 2000US-0216880P.	PR	08-NOV-2000; 2000US-0246523P.
PR	11-JUL-2000; 2000US-0217487P.	PR	08-NOV-2000; 2000US-0246525P.
PR	11-JUL-2000; 2000US-0217496P.	PR	08-NOV-2000; 2000US-0246526P.
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000; 2000US-0246527P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246528P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246533P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225213P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246611P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246613P.
PR	14-AUG-2000; 2000US-0225268P.	PR	17-NOV-2000; 2000US-0249207P.
PR	14-AUG-2000; 2000US-0225270P.	PR	17-NOV-2000; 2000US-0249208P.
PR	14-AUG-2000; 2000US-0225447P.	PR	17-NOV-2000; 2000US-0249209P.
PR	14-AUG-2000; 2000US-0225577P.	PR	17-NOV-2000; 2000US-0249210P.
PR	14-AUG-2000; 2000US-0225758P.	PR	17-NOV-2000; 2000US-0249211P.
PR	14-AUG-2000; 2000US-0225759P.	PR	17-NOV-2000; 2000US-0249212P.
PR	18-AUG-2000; 2000US-0226279P.	PR	17-NOV-2000; 2000US-0249213P.
PR	22-AUG-2000; 2000US-0226661P.	PR	17-NOV-2000; 2000US-0249214P.
PR	22-AUG-2000; 2000US-0226868P.	PR	17-NOV-2000; 2000US-0249215P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249216P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249217P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249218P.
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000; 2000US-0249219P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249244P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249245P.
PR	05-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249254P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249265P.
PR	06-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249297P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249299P.
PR	06-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249300P.
PR	08-SEP-2000; 2000US-0231242P.	PR	01-DEC-2000; 2000US-0250160P.
PR	08-SEP-2000; 2000US-0231243P.	PR	05-DEC-2000; 2000US-0251030P.
PR	08-SEP-2000; 2000US-0231244P.	PR	05-DEC-2000; 2000US-0251988P.
PR	08-SEP-2000; 2000US-0231413P.	PR	05-DEC-2000; 2000US-0256719P.
PR	08-SEP-2000; 2000US-0231414P.	PR	06-DEC-2000; 2000US-0251479P.
PR	08-SEP-2000; 2000US-0232080P.	PR	08-DEC-2000; 2000US-0251856P.
PR	08-SEP-2000; 2000US-0232081P.	PR	08-DEC-2000; 2000US-0251868P.
PR	12-SEP-2000; 2000US-0231968P.	PR	08-DEC-2000; 2000US-0251869P.
PR	14-SEP-2000; 2000US-0232397P.	PR	08-DEC-2000; 2000US-0251989P.
PR	14-SEP-2000; 2000US-0232398P.	PR	08-DEC-2000; 2000US-0251990P.
PR	14-SEP-2000; 2000US-0232399P.	PR	11-DEC-2000; 2000US-0254097P.
PR	14-SEP-2000; 2000US-0232400P.		
PR	14-SEP-2000; 2000US-0232401P.		
PR	14-SEP-2000; 2000US-0233063P.		
PR	14-SEP-2000; 2000US-0233064P.		
PR	14-SEP-2000; 2000US-0233065P.		
PR	21-SEP-2000; 2000US-0234223P.		
PR	21-SEP-2000; 2000US-0234274P.		

```
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
PR WPI; 2001-488785/53.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders.
XX
PS Disclosure; SEQ ID NO 317; 520pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of ovarian and breast antigens. These are shown in AAI62467-
CC AAI62572 and AAI62573. The sequences can be used in the
CC diagnosis, prevention and treatment of breast and ovarian cancers, and
CC their metastases. The present sequence is a genomic sequence of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at fp.wipo.int/pub/published\_pct\_sequences
XX
SQ Sequence 21441 BP; 5706 A; 4589 C; 4763 G; 6383 T; 0 U; 0 Other;
Query Match 7.3%; Score 36.8; DB 4; Length 21441;
Best Local Similarity 51.2%; Pred. No. 3.9;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 239 ATGCTCGGAAATTTCTCGAATATGTTTATTGAAATCACCAGATCTACGAGCCGACC 298
Db 8772 AGCTGTGTAATGGCTGTGCACACTATATTGCTTAGCACACATTCAGTGCAGACACAACC 8713
QY 299 ACTACTTTTCAGGCAATCCAGATAGCCCTGAGTTGGATTGCGCCCTTGGTTAAAGCAA 358
Db 8712 ATTGCTTTTTCTCAAAATATTTAAATCTGCAATTTGTTGAAGCCATGATGTGCACAGA 8653
QY 359 ATTCTTAGCCACAAAGATCTCGTCCGTCGCAATTTCTGAATTAGCT 406
Db 8652 ACCCAGATATGAGGACCGATTGTAAATGCAAAAGCGGAATTTCTCT 8605
RESULT 9
ID AAL06761/c
ID AAL06761 standard; DNA; 21441 BP.
XX
AC AAL06761;
XX
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9449.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-022513P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
```

20-OCT-2000; 2000US-0241808P.
20-OCT-2000; 2000US-0241809P.
20-OCT-2000; 2000US-0241826P.
01-NOV-2000; 2000US-0244617P.
01-NOV-2000; 2000US-0244617P.
08-NOV-2000; 2000US-0246474P.
08-NOV-2000; 2000US-0246475P.
08-NOV-2000; 2000US-0246476P.
08-NOV-2000; 2000US-0246477P.
08-NOV-2000; 2000US-0246478P.
08-NOV-2000; 2000US-0246479P.
08-NOV-2000; 2000US-0246523P.
08-NOV-2000; 2000US-0246524P.
08-NOV-2000; 2000US-0246525P.
08-NOV-2000; 2000US-0246526P.
08-NOV-2000; 2000US-0246527P.
08-NOV-2000; 2000US-0246528P.
08-NOV-2000; 2000US-0246532P.
08-NOV-2000; 2000US-0246609P.
08-NOV-2000; 2000US-0246610P.
08-NOV-2000; 2000US-0246611P.
08-NOV-2000; 2000US-0246613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
17-NOV-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465570/50.
Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.
Disclosure; SEQ ID NO 9449; 1297pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention
SQ Sequence 21441 BP; 5706 A; 4589 C; 4763 G; 6383 T; 0 U; 0 Other;
Query Match 7.3%; Score 36.8; DB 4; Length 21441;

Best Local Similarity 51.2%; Pred. No. 3.9;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 239 ATGGTCGGAATTTTCTCGAATATGTTATTGAAAATCACCAGAACTTACGAAGAGCCAGCC 298
DB 8772 AGGCTGTGTAATGGCTGTGACACTATATTGCTTAGCACACATTTCAGTGCAGACACACC 8713
QY 299 ACTACTTTTTCAGGCATTCCAGATAGCCCTGAGTGGATTGCGAGCTTTTGGTAAAGCAA 358
DB 8712 ATTGCTTTTTCCTCAAAATATTAAAATCTGCAATTGTTGAAGCCATGATGATGGACAGA 8653
QY 359 ATTCTTAGCCACAAGAAATCTCTGTCGTCGCAAAATCTGAATTAGCT 406
DB 8652 ACCACAGATATGAAGGACCGAATTGTAATGCAAAAGCGGAATCTCT 8605
RESULT 10
ACH24966/C
ID ACH24966 standard; cDNA; 458 BP.
XX
AC ACH24966;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult ovary cDNA #3346.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST; genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
PS WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 12178; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030073623

```
XX SQ Sequence 458 BP; 117 A; 81 C; 75 G; 178 T; 0 U; 7 Other;
Query Match 7.1%; Score 35.8; DB 9; Length 458;
Best Local Similarity 47.9%; Pred. No. 1.8;
Matches 103; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 179 TTAAGCCGAAGATGGCAAACTTTGCAAGTTTTCAGATCTTTTCAGATGCTTGCAAGC 238
DB 423 TAAACCCGAGGTAAAGTAATTTCTGAGAAATTTTCGACACAGAGAGTTAGCAACAC 364
QY 239 ATGTCGGAATTTTCTCGAATATGTTATTGAAATCACCAGATCTACGAGAGCCAGCC 298
DB 363 TTGATCAAAAATTTCCCTGATTTTGGAAAGGAAAAACAGGTTCCACAAAGTCAGATT 304
QY 299 ACTACTTTTCAGGATTCAGATAGCCCTGAGTTGGATTGAGCCTTTGTTGTTAAAGCAA 358
DB 303 AGTAAGACTGACGATTGACGAATTAAGTTAAATGAAAAATCAGACACATGATTGGAAGAT 244
QY 359 ATTCTTAGCCACAAAGATCTCTGTCCTGTCGCAAA 393
DB 243 GTTTGGAAGAAAAAATAATATCAGCCAGGCGCATA 209

RESULT 11
ADQ97960 0
WP Sequence split into 4 fragments LOCUS ADQ97960 Accession Adq97960
WP Fragment Name Begin End
WP ADQ97960_0 1 110000
WP ADQ97960_1 100001 210000
WP ADQ97960_2 200001 310000
WP ADQ97960_3 300001 390183
ID ADQ97960 standard; DNA; 390183 BP.
XX AC
XX ADQ97960;
XX DT
XX 07-OCT-2004 (first entry)
XX DE
XX Human cancer associated sequence HD11-029, SEQ ID 937.
XX DE
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX KW
XX Homo sapiens.
XX OS
XX W02004060304-A2.
XX PN
XX 22-JUL-2004.
XX PD
XX 22-DEC-2003; 2003WO-US041389.
XX PF
XX 27-DEC-2002; 2002US-00330773.
XX PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX PA
XX Morris DW, Malandro MS;
XX PI
XX WPI; 2004-543781/52.
XX DR
XX New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX PS
XX Claim 1; SEQ ID NO 937; 199pp; English.
XX CC
XX The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 390183 BP; 107624A; 72906C; 76848G; 123530T; 0U; 92750Other;
Query Match 7.1%; Score 35.6; DB 12; Length 110000;
Best Local Similarity 50.6%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 70 TGGCTTGATTTCTATGGGAGCCATATCTCTGGAATCCCAGAGATATGTTAGCTATGCA 129
DB 51069 TGGCACATCTGGTTGAAAACCTTAAACCAATTTGATTTTAAAAATATTGCATATATCTG 51128
QY 130 AGATTGGGGAAAATACCGGATGAATCACAGAGAGGCTCCACACTGTTTTTAAAGCCGAA 189
DB 51129 AGATTGGTAACTGTTAGAGATTAACAAAAAACAACACTTGCATAGTACTATATAAAA 51188
QY 190 GATGGCAAAACTTTGGCAAGTTTTCAGATCTTTTCAGATGCTTGCAACAGCA 239
DB 51189 TACACTATAACATAGCAAGCAATGCATTCATCAGCTGGCTTCACAAACA 51238

RESULT 12
ADF69148
ID ADF69148 standard; cDNA; 3160 BP.
XX AC
XX ADF69148;
XX DT
XX 12-FEB-2004 (first entry)
XX DE
XX Human MP53 nucleotide sequence SEQ ID NO:6.
XX KW
XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
XX KW Gene therapy; cancer; human; gene; ss.
XX OS
XX Homo sapiens.
XX PN
XX W02003083047-A2.
XX PD
XX 09-OCT-2003.
XX PF
XX 28-FEB-2003; 2003WO-US006025.
XX PR
XX 01-MAR-2002; 2002US-0361196P.
XX PA
XX (EXEL-) EXELIXIS INC.
XX KW
XX Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
XX PI Funke RP;
XX XX
XX WPI; 2003-812540/76.
XX DR P-PSDB; ADF69092.
XX CC
XX Identifying a candidate p53 pathway modulating agent for treating e.g.,
XX PT cancer by contacting an assay system comprising a MP53 polypeptide or
XX PT nucleic acid with a test agent and detecting a test agent-biased
XX PT activity.
XX XX
XX Example; SEQ ID NO 6; 406pp; English.
XX PS
XX The present invention describes a method for identifying a candidate p53
XX CC pathway modulating agent, which comprises: (a) providing an assay system
XX CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
XX CC fragment or derivative; (b) contacting the assay system with a test agent
XX CC under conditions where the system provides a reference activity except in
XX CC the presence of the test agent; and (c) detecting a test agent-biased
XX CC activity, where a difference between the test agent-biased activity and
XX CC the reference activity identifies the test agent as a candidate p53
XX CC pathway modulating agent. Also described: (1) modulating the p53 pathway
XX CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
XX CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
XX CC be used in gene therapy. The method is useful for identifying a candidate
XX CC p53 pathway modulating agent for preparing a composition for diagnosing
XX CC or treating e.g., cancer. The present sequence encodes a human MP53
XX CC protein, which is used in the exemplification of the present invention.
XX CC
XX SQ Sequence 3160 BP; 1089 A; 555 C; 523 G; 993 T; 0 U; 0 Other;
```



```
QY      291 AGCAGCCACTACTTTTCAGGCATTCAGCATAGCCCTGAGTTGGATTGCAGCCTTGTTGG 350
Db      69247 ATCAATCCAGCACCAAGAAGATTTTATATATGTCAAAATATGTCAATTAACAATAATTGT 69188

QY      351 TAAAGCAAAATCTTTAGCCACAAAGAAATCTCTGGTCCGTGCAGAAATCTTGAATTAGCTTTCA 410
Db      69187 TTCTCTCATCCATAACTATACAACTTTTAAAGTAGTGACCTTGTTGTTTTTATATTTA 69128

QY      411 AATGCTCTA 419
Db      69127 AATTCCTTA 69119

RESULT 15
AAS32838
ID      AAS32838 standard; DNA; 10113 BP.
XX
AC      AAS32838;
XX
DT      17-DEC-2001 (first entry)
XX
DE      Human genomic DNA for novel endocrine antigen, SEQ ID No 792.
XX
KW      Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW      thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW      antihense-therapy; antibody; endocrine disorder; hormone imbalance;
KW      reproductive disorder; endocrine cancer; pancreatic disorder;
KW      diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW      hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS      Homo sapiens.
XX
FN      WO200155319-A2.
XX
PD      02-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US001335.
XX
PR      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      14-AUG-2000; 2000US-0225759P.
PR      22-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226868P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231142P.
PR      08-SEP-2000; 2000US-0231143P.
PR      08-SEP-2000; 2000US-0231144P.
PR      08-SEP-2000; 2000US-0231143P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      21-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      23-SEP-2000; 2000US-0234597P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241121P.
PR      20-OCT-2000; 2000US-02411785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:04:38 ; Search time 114.576 Seconds
(without alignments)
7154.840 Million cell updates/sec

Title: US-10-018-929C-2_COPY_3000_3500
Perfect score: 501
Sequence: 1 agtcataactccagtcacat.....acacaaggttgtagtgcga 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.8	7.3	601	4	US-09-949-016-199237
C 2	36.8	7.3	1141	4	US-09-806-708B-22, Appli
C 3	36.8	7.3	134434	4	US-09-949-016-17362
C 4	34.6	6.9	1664976	4	US-08-916-421B-1
C 5	34.6	6.9	1664976	4	US-09-692-570-1
C 6	34.2	6.8	601	4	US-09-949-016-18847
C 7	34.2	6.8	601	4	US-09-949-016-18847
C 8	34.2	6.8	225127	4	US-09-949-016-16480
C 9	33.4	6.7	53451	4	US-09-949-016-18859
C 10	33	6.6	601	4	US-09-949-016-18848
C 11	33	6.6	601	4	US-09-949-016-23334
C 12	33	6.6	601	4	US-09-949-016-45601
C 13	33	6.6	601	4	US-09-949-016-56713
C 14	33	6.6	26075	4	US-09-949-016-12004
C 15	33	6.6	26076	4	US-09-949-016-13041
C 16	33	6.6	636591	4	US-09-949-016-11808
C 17	33	6.6	636591	4	US-09-949-016-13388
C 18	32.6	6.5	744	4	US-10-012-143-4
C 19	32.2	6.4	157644	4	US-09-949-016-16179
C 20	32.2	6.4	157644	4	US-09-949-016-16180
C 21	32	6.4	601	4	US-09-949-016-184492
C 22	32	6.4	38961	4	US-09-949-016-12143
C 23	32	6.4	38961	4	US-09-949-016-14474
C 24	32	6.4	76810	4	US-09-949-016-12528
C 25	32	6.4	78810	4	US-09-949-016-16198
C 26	32	6.4	190078	4	US-09-949-016-12707
C 27	32	6.4	190078	4	US-09-949-016-17026

28	31.8	6.3	285	4	US-09-513-999C-30335	Sequence 30335, A
29	31.8	6.3	1464	4	US-09-328-352-3247	Sequence 3247, Ap
c 30	31.8	6.3	124480	4	US-09-949-016-15921	Sequence 15921, A
31	31.6	6.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
32	31.4	6.3	801	4	US-09-543-681A-137	Sequence 137, App
c 33	31.4	6.3	1233	4	US-09-543-681A-3933	Sequence 3933, Ap
c 34	31.4	6.3	2169	3	US-09-434-408-3	Sequence 3, Appli
35	31.4	6.3	34047	4	US-09-949-016-13503	Sequence 13503, A
36	31.4	6.3	154600	4	US-09-949-016-14757	Sequence 14757, A
37	31.4	6.3	524032	4	US-09-949-016-16928	Sequence 16928, A
38	31.4	6.3	524032	4	US-09-949-016-16929	Sequence 16929, A
39	31.4	6.3	524032	4	US-09-949-016-16930	Sequence 16930, A
40	31.4	6.3	524032	4	US-09-949-016-16931	Sequence 16931, A
41	31.4	6.3	529885	4	US-09-949-016-14340	Sequence 14340, A
42	31.4	6.3	529885	4	US-09-949-016-14341	Sequence 14341, A
43	31.4	6.3	529885	4	US-09-949-016-14342	Sequence 14342, A
44	31.4	6.3	529885	4	US-09-949-016-14343	Sequence 14343, A
45	31.4	6.3	529885	4	US-09-949-016-14344	Sequence 14344, A

ALIGNMENTS

RESULT 1

US-09-949-016-199237/c
; Sequence 199237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09949.016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR APPLICATION NUMBER: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199237
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-199237

Query Match 7.3%; Score 36.8; DB 4; Length 601;
Best Local Similarity 51.2%; Pred. No. 0.13;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy	239	ATGTCGCGAATTTTCGATATGTTTAAATACCGAATCTACGAAGAGCCGCC	298
Db	170	AGGCTGTAAATGGCTGTGACATATATCTTAGCACATTCAGTCGACACACACC	111
Qy	299	ACTACTTTTCAGGCATTCACATAGCCCTGAGTTGGATTCAGGCTTTGTTGGTAAAGCAA	358
Db	110	ATTGCTTTTCTCAATATTTAAATCTGCAATTTGGTTGAAGCCATCGATGTGACACAGA	51
Qy	359	ATTCTTAGCCACAAGAAATCTCTGGTCGGTCCAAATTCCTGAATTAGCT	406
Db	50	ACCCACAGATATGAAGGCCGATTTGTAATGCAAAAGCGGAATCTCT	3

RESULT 2

US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741

```

; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match          7.3%; Score 36.8; DB 4; Length 1141;
Best Local Similarity 9.1%; Pred. No. 0.18;
Matches 45; Conservative 213; Mismatches 237; Indels 1; Gaps 1;

QY 5 ATAACTCCAGTCATCATGTAATCTTCTGGTACAGATGGTACATTTGGATGGAAATGAT 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 ADSWVWVWYANWRCRDYTYRNNYTKSYAHSYVWNSNNAMVYRRYSARNWSSMARWTR 540
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 GCTTTTGGCTTGATTCTATGGCGAGCCATATCTCTGGAATCCAGAGGATATGTTAGCT 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 NNWMSGBVRWRWAGTWWRHNNNNNDTRYWVWKKWARBTTVYDSMCKNAKSMWRGNN 480

QY 125 AGTCAAGATTGGGGGAAAAATACCGGATGAATCAAGAGGAGGCTCCACACTGTTTTAAAG 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 479 WRAMKMWMAANDAGAMDHTVYMGNNNTWRRRAKWNMAWCRRAVCNNNNNRACVWH 420

QY 185 CCGAAGATGCCAA-AACTTTGCCAAGTTTTCATCTTTCAGATGCTTGCAAGCATGGT 243
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 KHRWRTWTKYMKWTNNNNNNKAMVNRVAMVMSYRDTTNDMMMTSDWBWHVYVDYT 360

QY 244 CGGAAATTTTCTCGAATATGTTATTGAAATCACCGAATCTACGAGCAGCCACTAC 303
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 MRAWNNNNNNNRBCKTTSWMMWMDHNTHCTYGNNTWSAVBMAAMSWRAAGASNEVT 300

QY 304 TTTTCAGGCATTCAGATAGCCCTCAGTTCGATTCAGCCTTGTGTGTAAGCAAAATCT 363
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 YNWCWRMTYMGKWTNNNNNNKAWYRTKTVAWCNRNYYVDTAVWTBKNVYKYCYABWY 240

QY 364 TAGCCACAAAGATCTCTGTCGCGTCAAAATCTCGAATTAGCTTCAAAATGCTTAGAGT 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 YBMYMGKHWWRRRRASNNWMMVKCRNKYVSWHYHAMRYBKWABVGCNNNNWKDM 180

QY 424 AGAGGTGGATTATATTATTTCGATATTGCTCATGAAGAGTCTGTTCTCTGGAGCATAC 483
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 AHHHCATNNNNMMWWTAYVHHHMKKGAATWNNKTABRDDHBAHVITYWYDYWCAM 120

QY 484 ACAAGGTTTCAGTTC 499
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 CWNNAKAKVRTAMKHM 104

RESULT 3
US-09-949-016-17362/c
; Sequence 17362, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

```

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17362
; LENGTH: 134434
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-17362

Query Match          7.3%; Score 36.8; DB 4; Length 134434;
Best Local Similarity 51.2%; Pred. No. 1.6;
Matches 86; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 239 ATGTCGCGAAATTTCTCGAATATGTTATTTGAAATACCGGAATCTACGAAGAGCCAGCC 298
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119765 AGCTGTGTAATGGCTGTGACACTATATTTCTTAGCACACATTCAGTGCAGACACACC 119706

QY 299 ACTACTTTTTCAGGCATTCAGATAGCCCTGAGTTGGATTGCAGCCTTGTGTTAAAGCAA 358
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119705 ATTGCTTTTCTCAAAATATTTAAATCTGCAATTTGTTGAAGCCATGGATGTGGACAGA 119646

QY 359 ATTCTTAGCACAAAGAACTCTGGTCCGTCGGTGAATTTCTGAATTAGCT 406
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119645 ACCCAGACATATGAGGACCGATTGTAATGCAAAAGCGGAATTTCTCT 119598

RESULT 4
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; TITLE OF INVENTION: jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)

```

```
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (98343)..(98343)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (103998)..(103998)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (148948)..(148948)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (163385)..(163385)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (191989)..(191989)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (191995)..(191995)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (231980)..(231980)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (234187)..(234187)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (234220)..(234220)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (234814)..(234814)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (309398)..(309398)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (309418)..(309418)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (312837)..(312837)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (312993)..(312993)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (319226)..(319226)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (559167)..(559167)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (559241)..(559241)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (600992)..(600992)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (622708)..(622708)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (657081)..(657081)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (657203)..(657203)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (674435)..(674435)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (682442)..(682442)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (713652)..(713652)
;; OTHER INFORMATION: n equals a, t, c, or g
```

```
;; NAME/KEY: misc feature
;; LOCATION: (741684)..(741684)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (779455)..(779455)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (779676)..(779676)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (855539)..(855539)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (871619)..(871619)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1084830)..(1084830)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1096846)..(1096846)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1119881)..(1119881)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1130881)..(1130881)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1310988)..(1310988)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1313224)..(1313224)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1349473)..(1349473)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1349491)..(1349491)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1470091)..(1470091)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1569020)..(1569020)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1602912)..(1602912)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1603734)..(1603734)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1637998)..(1637998)
;; OTHER INFORMATION: n equals a, t, c, or g
;; NAME/KEY: misc feature
;; LOCATION: (1664854)..(1664855)
;; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

Query Match 6.9%; Score 34.6; DB 4; Length 1664976;
Best Local Similarity 46.2%; Pred. No. 27;
Matches 115; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```
Qy 171 ACACTGTTTAAAGCCGAGATGCGAAACCTTGGCCAACTTTGGCATCTTTTCAGATCCTT 230
Db 369367 ATACTCCCAATAGCCCACTTAGATAAACAATTGCTATGATGCATTAACAATAGTGT 369308
Qy 231 GCACAAGCATGGTCGGAATTTTCTCGAATATGTTATGAAATCACCGAATCTTACGAAG 290
Db 369307 AAACCAAGTATTAACAACTTACCTTCGAGAGGCAACTTAAACAATACTGCAACTAACAA 369248
Qy 291 AGCCAGCCACTACTTTTCAGGCATTCCAGATAGCCCTTCAGATTCGACCTTGTGG 350
```

Db 369247 ATCAATCCAGCACCAAGAAAGATTTTATATATGTCAAATATGTCATTAAACAATAATTGT 369188
QY 351 TAAAGCAATCTTAGCCACAAGAAATCTCTGGTCCGTGCAAAATTCGAATTAGCTTTCA 410
Db 369187 TTCTCTCATCCATAACTATCACAACTTTTTTAAAGTAGTGACCTTGRTTTTATTATTTA 369128
QY 411 AATGCTCTA 419
Db 369127 AATTCITTA 369119

RESULT 5
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Buit et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:

NAME/KEY: misc feature
LOCATION: (657081)..(657081)
FEATURE:
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (79676)..(79676)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g

Query Match
Best Local Similarity 46.2%; Pred. No. 27;
Matches 115; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
171 ACACGTGTTTAAAGCCGAAGATGCGAAACTTTCAGATCTTTCAGATGCTT 230

Db 369367 ATACTCCCAATAGCCAACTTAGATAAAACANTCTATGATGTCAATTAACATAGTGT 369308
Qy 231 GCACAAGCATGGTCGAAATTTTCTCGAATATGTTTATGAAATCACCGAATCTACGAAG 290
Db 369307 AAACAGTATTAAAACTACCTTCGAGAGGCAACTTTAACAATAACTGCAACTAACAA 369248
Qy 291 AGCAGGCACTACTTTTCAGGCATTCAGATAGCCCTGAGTGGATTCGACCCCTTGTGG 350
Db 369247 ATCAATCCAGCACCAGAAAGATTTATATATGTCAAAATATTGCAATTAACAAATATGT 369188
Qy 351 TAAAGCAAAATTTCTAGCCACAAGAATCTCTGGTCGCAAAATCTGAAATTAGCTTCA 410
Db 369187 TTCTCTCATCCATCACTATCACAACTTTTAAAGTAGTAGACCTTGCTGTTTATTTA 369128
Qy 411 AATGCTCTA 419
Db 369127 AATCTTTA 369119

RESULT 6

US-09-949-016-18847/c
; Sequence 18847, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18847
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18847

Query Match 6.8%; Score 34.2; DB 4; Length 601;
Best Local Similarity 63.0%; Pred. No. 0.9;
Matches 51; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

Qy 202 TTGCCAAGTTTTCATCTTTCAGATGTTGCAAGCATGTCGAAATTTTCGGAATA 261
Db 302 TYGCTATGGCTGCATTTTTCAGTTCCTTCTCAGGAGTGATGTGTTGTTGATC 243
Qy 262 TGTATTGAAATCACCAGAT 282
Db 242 AGTGATTCAAAATTACCAAT 222

RESULT 7

US-09-949-016-56712/c
; Sequence 56712, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56712
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56712

Query Match          6.8%; Score 34.2; DB 4; Length 601;
Best Local Similarity 63.0%; Pred. No. 0.9;
Matches 51; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

Qy 202 TTGCCAAGTTTTCAGTCTTTCAGATGCTTGCACAGCATGGTCGGAATTTTCTCGAATA 261
Db 302 TYGCTATGGCTGCATTTTTCAGTCTTTCAGTCTTTCAGGAGTGTAGTAGATTTTGTGATC 243

Qy 262 TGTATTGAAATCACCAGAT 282
Db 242 AGTATTACAAATACCAGAT 222

RESULT 8
US-09-949-016-16480
; Sequence 16480, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16480
; LENGTH: 225127
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(225127)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16480

Query Match          6.8%; Score 34.2; DB 4; Length 225127;
Best Local Similarity 47.1%; Pred. No. 14;
Matches 105; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 16 GTACATCATGTAATCTTCTGTACAGATGGTACATTTGGATGGAAATGCTTTTGGCTT 75
Db 85154 GTAATCAAGGAATATCTGTTATGCTGTAATATAAGGAATGAGAACTTGGGTG 85213

Qy 76 GTATTTATGGGAGCATATCTCTGGAATCCAGAGATATGTTAGCTAGTCAAGATTG 135
Db 85214 GAATCAAGGGCAACTTTTCCCTCGAATCCCTGAAGTCCCTAATATAGGAGATCT 85273

Qy 136 GGGGAAATACCGGATGAATCACAGAGAGCTCCACACTGTTTTTAAAGCGGAAGATGGC 195
Db 85274 CTAATAATGTTGGAGGGGTGCAAGGGGTAGGGGAGACAAAAAATAAATCTGAGGAATT 85333

Qy 196 ABACTTTGCCAGCTTTTGCATCTTTTCAGATGCTTGCACAGC 238
Db 85334 TCCACTTGACCAATATACACTTCTCATGTGTAGTTACAGATC 85376

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56712
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56712

Query Match          6.8%; Score 34.2; DB 4; Length 601;
Best Local Similarity 63.0%; Pred. No. 0.9;
Matches 51; Conservative 1; Mismatches 29; Indels 0; Gaps 0;

Qy 202 TTGCCAAGTTTTCAGTCTTTCAGATGCTTGCACAGCATGGTCGGAATTTTCTCGAATA 261
Db 302 TYGCTATGGCTGCATTTTTCAGTCTTTCAGTCTTTCAGGAGTGTAGTAGATTTTGTGATC 243

Qy 262 TGTATTGAAATCACCAGAT 282
Db 242 AGTATTACAAATACCAGAT 222

RESULT 9
US-09-949-016-15859
; Sequence 15859, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15859
; LENGTH: 53451
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15859

Query Match          6.7%; Score 33.4; DB 4; Length 53451;
Best Local Similarity 57.0%; Pred. No. 13;
Matches 61; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 265 TATTGAAATCACCAGATCTACGAGAGCCAGCCACTACTTTTCAGGCATTCAGATAGC 324
Db 28034 TATTTAACTCCCTGGATCTGTATCAAGGCTTCATTAGTTTTCATCCATTGCAATTTGC 28093

Qy 325 CTTGAGTTGGATTCAGCCCTTGTGTAAGCAATCTTAGCCACA 371
Db 28094 CTAGAGTTAGATAGCTGGTTAATGCAGGAATCCAGATTAAGCCAGA 28140

RESULT 10
US-09-949-016-18848/c
; Sequence 18848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18848
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18848

Query Match          6.6%; Score 33; DB 4; Length 601;
Best Local Similarity 63.0%; Pred. No. 2.2;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 202 TTGCCAAGTTTTCAGTCTTTCAGATGCTTGCACAGCATGGTCGGAATTTTCTCGAATA 261
Db 97 TCCTATGGCTGCATTTTTCAGTCTTTCAGTCTTTCAGGAGTGTAGTAGATTTTGTGATC 38

Qy 262 TGTATTGAAATCACCAGAT 282
```

Db 37 AGTGATTCACAAATTACCAAT 17

RESULT 11

US-09-949-016-23334
; Sequence 23334, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 23334
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-23334

Query Match 6.6%; Score 33; DB 4; Length 601;
Best Local Similarity 51.0%; Pred. No. 2.2;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 326 CTGAGTTGGATTGCGAGCCTTTGGTAAAGCAAAATCTTAGCCACAAAGAAATCTCTGGTC 385
Db 328 CTGGCTGGCTTTCTCAGGTGTGTGCGAGGCAATGAGGCGAGCTCTACAGAATTAAATCCAG 387
Qy 386 CGTGCAAATTTCTGAATTTAGCTTTCAATGCTCTAGAGTAGAGGTGGATTATTTATTCG 445
Db 388 AATCCATAATTCGGATATTTCTTAACAGCTTCAATGTATTGTTGAAATAGCCTCTCCCA 447
Qy 446 ATATTGCTCGCATGAAGAGTCTGTTCTCTGGAG 478
Db 448 AATGTACCCGTATTTGATTCCTTTATGTTG 480

RESULT 12

US-09-949-016-45601
; Sequence 45601, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45601
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-45601

Query Match 6.6%; Score 33; DB 4; Length 601;
Best Local Similarity 51.0%; Pred. No. 2.2;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 326 CTGAGTTGGATTGCGAGCCTTTGGTAAAGCAAAATCTTAGCCACAAAGAAATCTCTGGTC 385
Db 328 CTGGCTGGCTTTCTCAGGTGTGTGCGAGGCAATGAGGCGAGCTCTACAGAATTAAATCCAG 387
Qy 386 CGTGCAAATTTCTGAATTTAGCTTTCAATGCTCTAGAGTAGAGGTGGATTATTTATTCG 445
Db 388 AATCCATAATTCGGATATTTCTTAACAGCTTCAATGTATTGTTGAAATAGCCTCTCCCA 447
Qy 446 ATATTGCTCGCATGAAGAGTCTGTTCTCTGGAG 478
Db 448 AATGTACCCGTATTTGATTCCTTTATGTTG 480

RESULT 13

US-09-949-016-56713/c
; Sequence 56713, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56713
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56713

Query Match 6.6%; Score 33; DB 4; Length 601;
Best Local Similarity 63.0%; Pred. No. 2.2;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 202 TTGCCAAGTTTGCATCTTTTCAGATGCTTCACAGCATGCTGCAGAAATTTTCTCGAATA 261
Db 97 TCGCTATGGCTGCAATTTTTCAGTTCCTTCTCAGGAGTGATGTAGATTTTGTGATC 38
Qy 262 TGTATTGAAATCACCAGAT 282
Db 37 AGTGATTCACAAATTACCAAT 17

RESULT 14

US-09-949-016-12004
; Sequence 12004, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12004
; LENGTH: 26075

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12004

Query Match      6.6%; Score 33; DB 4; Length 26075;
Best Local Similarity 51.0%; Pred.No.13;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 326 CTGAGTTGGATTGCAGCCTTTGGTAAAGCAAAATCTTAGCCACAAAGAATCTCTGGTC 385
Db 18884 CTGGCTGGGCTTTCTCACGTGTGTGCAGGGAATGAGGCAGCTCTACAGAAATTAATCCAG 18943

QY 386 CGTGCAAATTCGAATTAGCTTTCAAAATGCTCTAGAGTAGAGGTGGATTATATTATTCG 445
Db 18944 AATCCATATTCGGATATTCTTTAACAGCTTCAAAATGATATTGTGAATAGCCTCTCCCA 19003

QY 446 ATATTGTCCTGCATGAAGAGTCTGTTCCCTGGAG 478
Db 19004 AAATGTACCCGTATTTTGATTCCCTTTATGGTG 19036

RESULT 15
US-09-949-016-13041
; Sequence 13041, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13041
; LENGTH: 26076
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13041

Query Match      6.6%; Score 33; DB 4; Length 26076;
Best Local Similarity 51.0%; Pred.No.13;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 326 CTGAGTTGGATTGCAGCCTTTGGTAAAGCAAAATCTTAGCCACAAAGAATCTCTGGTC 385
Db 18884 CTGGCTGGGCTTTCTCACGTGTGTGCAGGGAATGAGGCAGCTCTACAGAAATTAATCCAG 18943

QY 386 CGTGCAAATTCGAATTAGCTTTCAAAATGCTCTAGAGTAGAGGTGGATTATATTATTCG 445
Db 18944 AATCCATATTCGGATATTCTTTAACAGCTTCAAAATGATATTGTGAATAGCCTCTCCCA 19003

QY 446 ATATTGTCCTGCATGAAGAGTCTGTTCCCTGGAG 478
Db 19004 AAATGTACCCGTATTTTGATTCCCTTTATGGTG 19036
```

Search completed: July 12, 2005, 10:18:09
Job time : 125.576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 543.528 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929c-2_COPY_3000_3500

Perfect score: 501

Sequence: 1 agtcataactccagtcacat.....acacaaagtttcagttcga 501

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 13661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82.6	16.5	6409	US-10-437-963-76223	Sequence 76223, A
2	55.4	11.1	661	US-10-425-115-145108	Sequence 145108, A
3	53.2	10.6	487	US-10-424-599-12966	Sequence 12966, A
4	45.6	9.1	8425	US-10-437-963-69891	Sequence 69891, A
5	44.6	8.9	845	US-10-424-599-28522	Sequence 28522, A
6	37.2	7.4	5865	US-10-437-963-76215	Sequence 76215, A
7	37	7.4	1255	US-10-424-599-95539	Sequence 95539, A

c	8	36.8	7.3	21441	10	US-09-764-891-9449	Sequence 9449, Ap
c	9	36.8	7.3	21441	17	US-10-091-414-317	Sequence 317, App
c	10	36.6	7.3	2940917	13	US-10-027-633-174763	Sequence 174763, A
c	11	36.6	7.3	2940917	17	US-10-027-633-174763	Sequence 174763, A
c	12	35.8	7.1	458	10	US-09-918-995-12178	Sequence 12178, A
c	13	35.6	7.1	561515	19	US-10-741-601-5682	Sequence 5682, Ap
c	14	35.6	7.1	561515	21	US-10-741-601-5682	Sequence 17730, A
c	15	35	7.0	401433	22	US-10-737-082-79	Sequence 79, Appl
c	16	35	7.0	401433	22	US-10-765-790-79	Sequence 79, Appl
c	17	34.8	6.9	534	19	US-10-021-323-10615	Sequence 10615, A
c	18	34.4	6.9	542	18	US-10-424-599-126300	Sequence 126300, A
c	19	34	6.8	495	13	US-10-027-632-94232	Sequence 94232, A
c	20	34	6.8	495	13	US-10-027-632-307917	Sequence 307917, A
c	21	34	6.8	495	17	US-10-027-632-94232	Sequence 94232, A
c	22	34	6.8	495	17	US-10-027-632-307917	Sequence 307917, A
c	23	34	6.8	10113	17	US-10-074-024-792	Sequence 792, App
c	24	34	6.8	11655	17	US-10-074-024-792	Sequence 792, App
c	25	33.6	6.7	88493	19	US-10-027-632-278628	Sequence 278628, A
c	26	33.2	6.6	499	13	US-10-027-632-278628	Sequence 278628, A
c	27	33.2	6.6	499	13	US-10-027-632-278628	Sequence 278628, A
c	28	33.2	6.6	499	17	US-10-027-632-278628	Sequence 278628, A
c	29	33.2	6.6	499	17	US-10-027-632-278628	Sequence 278628, A
c	30	33.2	6.6	937	18	US-10-425-114-18981	Sequence 18981, A
c	31	33.2	6.6	950	21	US-10-953-856-34	Sequence 34, Appl
c	32	33.2	6.6	1449	17	US-10-282-122A-8959	Sequence 8959, Ap
c	33	33.2	6.6	5879	15	US-10-311-455-242	Sequence 242, App
c	34	33	6.6	606	19	US-10-767-701-456	Sequence 456, App
c	35	33	6.6	1224	17	US-10-282-122A-32948	Sequence 32948, A
c	36	32.8	6.5	67832	11	US-09-997-722-67	Sequence 67, Appl
c	37	32.8	6.5	135827	19	US-10-322-281-232	Sequence 232, App
c	38	32.8	6.5	1223197	13	US-10-027-632-179264	Sequence 179264, A
c	39	32.8	6.5	1223197	17	US-10-027-632-179264	Sequence 179264, A
c	40	32.6	6.5	387	20	US-10-425-115-44733	Sequence 44733, A
c	41	32.6	6.5	469	14	US-10-060-036-1521	Sequence 1521, Ap
c	42	32.6	6.5	744	16	US-10-012-143-4	Sequence 4, Appl
c	43	32.6	6.5	2623	20	US-10-483-512-72	Sequence 72, Appl
c	44	32.6	6.5	2719	17	US-10-108-260A-1193	Sequence 1193, Ap
c	45	32.6	6.5	2750	15	US-10-050-882-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1

US-10-437-963-76223
; Sequence 76223, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76223
; LENGTH: 6409
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_76236C.1
US-10-437-963-76223

Query Match 16.5%; Score 82.6; DB 19; Length 6409;
Best Local Similarity 52.8%; Pred. No. 4e-15;
Matches 178; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy	135	GGGGGAAAATACCGGATGAATCAGAGGAGGCTCCACACTGTTTTAAAGCCGGAAGATGG	194
Db	3032	GGGCAGAAAGCCTTTGAGGGCACAGAAAAGGTCTTCACCCAACTCAAGCCAGAGCTGT	3091
Qy	195	CAAAACTTTGGCRAAGTTTTCATCTTTTCAGATGCTTTCACAAAGCATGGTCGGAATTTTC	254
Db	3092	CAAAAGTTGATGAATCTGTGGAAATTGGCGGAGACTGTCAAATGTTTATGTGAGGAATTTAC	3151
Qy	255	TCGAATATGTTATTGAAAATCACCCGAATCTACGAAAGAGCCAGGCACACTACTTTTCAGGCAT	314
Db	3152	TTGATTACATTTTGAAGAATCATCAAGTCAGTCAGGAGCCAAAGGCATATTGTCATGCAT	3211
Qy	315	TCCAGATAGCCCTCAGATTGGATTGCAGCCTTGTTGGTAAAGCAAATCTTTAGCCACAAG	374
Db	3212	TCACATTTGCCCTGTTTGGCGTGCTGCTCTCTTTAAGCATTAAGATCAATCGAAGAG	3271
Qy	375	AATCTCTGGTCGGTGCAAATTTCTGAATTTAGCTTTCAAATGCTCTAGAGTAGAGGTGGATT	434
Db	3272	AGTCACTTGCCCTTTCTGTGAGAAACTTGAACATATGATGTCGATGAAGTGTCTGCCTGAAT	3331
Qy	435	ATATTTATTGAAATGTTCTCGCATGAAGAGCTCTGTT	471
Db	3332	ATGTTTATGAAAGCTAAGGATCCTCAAGAAAAAGTT	3368

```

RESULT 2
US-10-425-115-145108
; Sequence 145108, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 145108
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_63821C.1
US-10-425-115-145108

```

RESULT 3
 US-10-424-599-12966
 ; Sequence 12966, Application US/10424599-
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 12966
 ; LENGTH: 487
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_111714C.1
 US-10-424-599-12966

RESULT 4
US-10-437-963-69891/c
; Sequence 69891, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules and Uses Thereof for Plant Improvement
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69891
; LENGTH: 8425
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70516C.1
US-10-437-963-69891

Query Match 9.1%; Score 45.6; DB 19; Length 8425;

;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 174763
;; LENGTH: 2940917
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(2940917)
;; OTHER INFORMATION: n = A,T,C or G
;; US-10-027-632-174763

Query Match 7.3%; Score 36.6; DB 17; Length 2940917;
Best Local Similarity 53.1%; Pred. No. 76;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 161 AGGAGCTCCACACTGTTTAAAGCCGAGATGCGCAAACTTTCGCAAGTTTTCATCTT 220
Db 1206194 AGAAACCTCTACATGTAGATAGTAACAGGCAATCGCAAAATTTTCTCAGGATTTCTGGAGA 1206135

Qy 221 TCAGATGCTGCACAGCATGTCGGAATTTTCTCGAATATGTTATGAAATCACCGA 280
Db 1206134 TGCATATCTTAGACATCTTACCTTTCTTCTCGGATTTTTCCTGATAATCTCTTT 1206075

Qy 281 ATCTACGAAGAGCCAGCCACTACTTTT 307
Db 1206074 ATCGACGTGCTACTAATTCGTTCTATT 1206048

RESULT 12
US-09-918-995-12178/c
;; Sequence 12178, Application US/09918995
;; Publication No. US20030073623A1
;; GENERAL INFORMATION:
;; APPLICANT: Hyseq, Inc.
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;; FROM VARIOUS CDNA LIBRARIES
;; FILE REFERENCE: 20411-756
;; CURRENT APPLICATION NUMBER: US/09/918,995
;; CURRENT FILING DATE: 2001-07-30
;; PRIOR APPLICATION NUMBER: US/09/235,076
;; PRIOR FILING DATE: 1999-01-20
;; NUMBER OF SEQ ID NOS: 38054
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 12178
;; LENGTH: 458
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)...(458)
;; OTHER INFORMATION: n = A,T,C or G
;; US-09-918-995-12178

Query Match 7.1%; Score 35.8; DB 10; Length 458;
Best Local Similarity 47.9%; Pred. No. 1.2;
Matches 103; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 179 TTAAGCGGAGATGGCAAACTTTGCAAGTTTTCATCTTTTCAGATGCTTGCACAGC 238
Db 423 TAAAAACCCGAGGTAAGTAATTTCTGAGAAATTTTTCGACAAACAGAGAGTTAGCAACAC 364

Qy 239 ATGTCGGAATTTTCTCGAATATGTTTATTGAAATCACCAGAACTTACGAAGAGCCAGCC 298
Db 363 TTGATCAAAAATTTCCCTGATTTTGGAAAGGAAAAAAGGTTCCACNAAGTCAGATT 304

Qy 299 ACTACTTTTTCAGGCATTCAGATAGCCCTGAGTTGGATGTCAGCTTTGTTGGTAAAGCAA 358
Db 303 AGTAAGACTGACGATTGACGAATAAAGTTAAATGAAAAATCAGACACATGATTTGAAGAT 244

Qy 359 ATTCTTAGCCACAAAGATCTCTGTCGTCGCAAA 393
Db 243 GTTTGGAAGAAAAAATAATCAGCCAGGGCATA 209

RESULT 13
US-10-741-601-5682
;; Sequence 5682, Application US/10741601
;; Publication No. US20040166519A1
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele et al.
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; STENOSIS, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001500
;; CURRENT APPLICATION NUMBER: US/10/741,601
;; CURRENT FILING DATE: 2003-12-22
;; NUMBER OF SEQ ID NOS: 26415
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5682
;; LENGTH: 561515
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)...(561515)
;; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5682

Query Match 7.1%; Score 35.6; DB 19; Length 561515;
Best Local Similarity 51.2%; Pred. No. 67;
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 260 TATGTTATTGAAATCACCAGAACTTACGAAGAGCCAGCCACTACTTTTCAGGCATTCAG 319
Db 230330 TAGATTCAGTAATATTATCAAGTCTCATATAAAACAGGTTGMAATTTTCTACTGAATACAG 230389

Qy 320 ATAGCCCTGAGTTGGATTGAGCCTTTGTTGTAAGCAAAATCTTAGCCACAAAGAATCT 379
Db 230390 ATAACTCTGGAATGAATGACACCTTTATGTAGTAGTTTCTGCTCCCAATCTCTGAAC 230449

Qy 380 CTGTCGTCGCAAAATCTGAATTAGCTTCAAAATGCTCTAGA 421
Db 230450 ATGTCGTCAGTTTATTAGATTTTATTATATCTTTTAAA 230491

RESULT 14
US-10-741-600-17730
;; Sequence 17730, Application US/10741600
;; Publication No. US20050026169A1
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele et al.
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001499
;; CURRENT APPLICATION NUMBER: US/10/741,600
;; CURRENT FILING DATE: 2003-12-22
;; NUMBER OF SEQ ID NOS: 73997
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17730
;; LENGTH: 561515
;; TYPE: DNA

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(561515)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17730

Query Match          7.1%; Score 35.6; DB 21; Length 561515;
Best Local Similarity 51.2%; Pred. No. 67;
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 260 TATGTTATTGAAATCACCAGATCTAGCAGAGCCAGCCACTCTTTTCAGGCAATCCAG 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230330 TAGATTCAGTAATATTCAAGTCTCATAAAACCAGTTGARATTTTCACTGAATACAG 230389

QY 320 ATAGCCCTGAGTGGATTGAGCCCTTGTGTTAAAGCAAAATCTTAGCCACAAAGAAATCT 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230390 ATAGTCTGGAATGAATTGACACCTTTATGTTAGTGAGTTTCTGCCCAATTCTCTGAAC 230449

QY 380 CTGTCGTCGCAAAATCTGAATTAGCTTTCAAATGCTCTAGA 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230450 ATGTCAGTATTATTAGATTTTATTATTAATCTTTAAA 230491

RESULT 15
US-10-737-082-79
; Sequence 79, Application US/10737082
; Publication No. US20050130170A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2032
; CURRENT APPLICATION NUMBER: US/10/737,082
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79
; LENGTH: 401433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-737-082-79

Query Match          7.0%; Score 35; DB 22; Length 401433;
Best Local Similarity 57.9%; Pred. No. 87;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 187 GAAGATGCCAAAACCTTTGCCAAGTTTTCAGATCTTTGACATGCTTGCAAGCATGTCGG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25469 GAAGAAGGGAGGAGCATTCGCCAAGAAAGTCATATTGAGCTGGATATTGAAGAATGAGTTG 25528

QY 247 AAATTTTCTCGAATATGTTATTGAAATCACCAGAAATCTACGAAGAGC 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25529 AAATGTTCTAGGAGAGAAATTTGGAAGGAATGCGTCATCCGAGAGC 25575

Search completed: July 12, 2005, 11:12:43
Job time : 556.528 secs
```



```

Qy 330 GTTCGATTGCAGCTTGTGGTAAAGCAAAATCTTAGCCACAAAGAAATCTCTGTCGCTG 389
Db 365 TCAGGGCTGCTGCTTCTTGTCTTAAATCACAAGTTGACCAAGAACTATTCTGCTCTTT 424
Qy 390 CAAATTCGTAATTAGCTTTCAAATGCTCTAGAGTAGAGGTGGATTATATTATTTCGATAT 449
Db 425 CAAAGAACAAATGAAATTTTCACTGTAAGCAGCAGAGGTGGACTATAGTATTCTTTTC 484
Qy 450 TGTCTCTGATGAAGAGTCTGTTCCTCGGAGCATACA 484
Db 485 TGCAGTGTTCGAGAAATCTTTGTATATCGTACA 519

RESULT 4
AU288479
LOCUS AU288479 zinnia cultured mesophyll cell equalized cDNA Zinnia EST 04-DEC-2002
DEFINITION AU288479 zinnia cultured mesophyll cell equalized cDNA Zinnia
ACCESSION AU288479
VERSION AU288479.1 GI:24248599
KEYWORDS EST.
SOURCE zinnia elegans
ORGANISM zinnia elegans
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Asteroideae;
Heliantheae; Zinnia.
REFERENCE 1 (bases 1 to 516)
AUTHORS Demura, T., Tashiro, G., Horiguchi, G., Kishimoto, N., Kubo, M.,
Matsuoka, N., Minami, A., Nagata-Hiwataashi, M., Nakamura, K.,
Okamura, Y., Sassa, N., Suzuki, S., Yazaki, J., Kikuchi, S. and
Fukuda, H.
TITLE Visualization by comprehensive microarray analysis of gene
expression programs during transdifferentiation of mesophyll cells
into xylem cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT Contact: Taku Demura
Morphogenesis Research Group
RIKEN Plant Science Center
1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9605
Fax: 81-45-503-9573
Email: demura@postman.riken.go.jp
This clone was obtained at our laboratory.
Seq primer: M13 forward.
FEATURES             source
1. 516
/organism="Zinnia elegans"
/mol_type="mRNA"
/cultivar="Canary bird"
/db_xref="taxon:34245"
/clone="Z2904"
/tissue_type="mesophyll cell"
/clone_lib="zinnia cultured mesophyll cell equalized cDNA"
/notes="Vector: pGEM-T easy; cultured in tracheary element
differentiation-inductive medium"
ORIGIN
Query Match 16.0%; Score 80.2; DB 1; Length 516;
Best Local Similarity 56.1%; Pred. No. 5e-13;
Matches 171; Conservative 0; Mismatches 133; Indels 1; Gaps 1;

Qy 180 TAAAGCCGAAGATGGCAAACTTTGCCAAGTTTTCATCTTTTCAGATGCTTTCACAAGCA 239
Db 110 TGAAGCTTAAATACCTAACTATGTGAAGTCTTAAAGTTCGCGAGGATGTGAATAA 169
Qy 240 TGGTCGGAATTTCTCGAATATGTTATGTAATACCCGAATCTACGAAGCCAGCCA 299
Db 170 TGGTGAAGGATTTCTCGAGTAGTGTATTAGAAATTATCGCGTCAACGAAGCACTGAAA 229
Qy 300 CTACTTTTCAGCATTCAGATAGCCCTGAGTTTGGATTGCGAGCTTGTTCGTAAGCAA 358
Db 230 GCACATTACATGCTTTCTTGATATCCCTGTGTGGATTGTTCCGGATTGTTAAACAC 289

```

```

Qy 359 ATTCTTAGCCACAAAGAAATCTCTGTCGTCGTCGCAAAATCTGAATTAGCTTTCAATGCTCT 418
Db 290 AAGCTTGTATAGGAAGCAATCATAAGTCTTCGCAACGAACACTTGGGTTTCAGTTGTAAG 349
Qy 419 AGAGTAGAGGTGGATTATATTATTTCGATATTTGCTTCGATGAGAGTCTCTTCTCTGAG 478
Db 350 GAAATGAGCAAACTCAGTGTACTCTGAAGCTGGAATTTGCAAAAGAGATGTTCTTACTT 409
Qy 479 CATAC 483
Db 410 CACAC 414

RESULT 5
BQ873535
LOCUS BQ873535 529 bp mRNA linear EST 15-AUG-2002
DEFINITION QG11M23.yg.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION BQ873535
VERSION BQ873535.1 GI:22260094
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 529)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolkmann, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Amundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QG11 row: M column: 23.
FEATURES             Location/Qualifiers
1. 529
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG11M23"
/lab_host="E.coli"
/clone_lib="QG-ABCDI lettuce salinas"
/notes="Vector: pBRcNAsf1AB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG TISSUE=flowers pre-fertilized
TAG LIB=QG ABCDI lettuce salinas
TAG_SEQ=GCCTGACGGG"
ORIGIN
Query Match 15.6%; Score 78; DB 5; Length 529;
Best Local Similarity 55.6%; Pred. No. 2.4e-12;
Matches 150; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 177 TTTTAAAGCCGAAGATGGCAAACTTTGCCAAGTTTTCATCTTTTCAGATGCTTCGACAA 236

```

Db	55	TTCTACATCCCAATGTGCATGAGCTTTGTAAATCCTAAAGTTTCCGAGGATGTCAGA	114
Qy	237	GCATGGTCGGAATTTTCTCGAAATATGTTATTGAAAAATCAGCAATCTACGAAGAGCCAG	296
Db	115	ACAAAGTGGATGGATTTCTTGATTTGTTCTCAAGAAATTAATGTCAGTAGGGAGGATA	174
Qy	297	CCACTACTTTTCAGGCATTCAGATAGCCCTGAGTTCGATTCGAGCCCTGTTGTGTAAGC	356
Db	175	AAACTACTCTACAGGCATTCATGATATCATTTGTGTGGATTGGGCTTCTCTTGCAAAAC	234
Qy	357	AAATTCCTTAGCCCAAGAAATCTCTGGTCCGTGCAAAATCTGAAATTAGCTTTCAAATGCT	416
Db	235	ATAAAATTGATAGAAAGAGGCTTTTGACCTTGCAAAAAAGCAATGGAATTTTGATTGTA	294
Qy	417	CTAGAGTAGAGGTGGATTAATTTATTTCGA	446
Db	295	AGGAAGAACAGGTGAATTTCTGTTCACTTGA	324
RESULT 6			
BH979933			
LOCUS	ode69f11.b1	B.oleracea002 Brassica oleracea genomic, genomic survey	712 bp DNA linear GSS 07-OCT-2002
DEFINITION	sequence.		
ACCESSION	BH979933		
VERSION	BH979933.1		
KEYWORDS	GI:23510550		
SOURCE	GSS.		
ORGANISM	Brassica oleracea		
REFERENCE	Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash, W., Rabinowicz, P.D. and Wilson, R.K.		
AUTHORS	Whole genome shotgun reads from Brassica oleracea		
TITLE	Unpublished (2002)		
JOURNAL	Contact: Richard K. Wilson		
COMMENT	Genome Sequencing Center Washington University School of Medicine Email: submissions@watson.wustl.edu Plate: ode69 row: f column: 11 Seq primer: -21UpPOT forward Class: shotgun High quality sequence start: 8 High quality sequence stop: 551.		
FEATURES	Location/Qualifiers		
source	1..712		
	/organism="Brassica oleracea"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3712"		
	/clone_lib="B.oleracea002"		
	/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T01000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."		
ORIGIN			
Query Match	12.9%	Score 64.4;	DB 8; Length 712;
Best Local Similarity	70.5%	Pred. No. 3.7e-08;	
Matches	86; Conservative	0; Mismatches 36; Indels	0; Gaps 0;
Qy	212	TTGCATCTTTTCAGATGCTTGCAAGCATGTCGGAATTTTCTCGAATATGTTATTGAA	271
Db	427	TTTTGTTTTCAGGAAGATTCGGCAGCATAGCTGAAAGTTTCTTGAATATATTGAA	486
Qy	272	AATCACCAGATCTACGAAGCCAGCCACTACTTTTCAGGCATTCAGATAGCCCTGAGT	331
Db	487	CACCACCGTGTATGCACAGACCCAGCCACCATAATTGCAGGCGTTCCAGATAGCCTTGTA	546

Qy 332 TG 333
Db 547 TG 548

RESULT 7

CV262174/c	342 bp	mRNA	linear	EST 22-SEP-2004
LOCUS	WS02018.B21_N09	PTxN-IB-N-A-11	Populus balsamifera subsp. trichocarpa x Populus nigra	cDNA clone WS02018_N09 3', mRNA sequence.
ACCESSION	CV262174			
VERSION	CV262174.1	GI:52515149		
KEYWORDS	EST.			
SOURCE	Populus balsamifera subsp. trichocarpa x Populus nigra			
ORGANISM	Populus balsamifera subsp. trichocarpa x Populus nigra			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustoside I; Malpighiales; Salicaceae; Populus.			
AUTHORS	1. (bases 1 to 342) Ralph, S., Cooper, D., Kolosova, N., Oddy, C., Butterfield, Y., Kirkpatrick, R., Liu, J., Palmquist, D., Stott, J., Barber, S., Yang, G., Babakiaff, R., Brown-John, M., Chand, S., Featherstone, R., Maeson, A., Mayo, M., Moran, J., Olson, T., Wong, D., Ritland, C.E., Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C., Ritland, K. and Bohlmann, J.			
TITLE	The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries			
JOURNAL	Unpublished (2004)			
COMMENT	Contact: Joerg Bohlmann Genome BC forest genomics program University of British Columbia UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3 Tel: 1-604-822-0282 Fax: 1-604-822-6097 Email: bohlmann@interchange.ubc.ca Plate: WS02018 row: N column: 09 High quality sequence stop: 342.			
FEATURES	Location/Qualifiers			
source	1..342			
	/organism="Populus balsamifera subsp. trichocarpa x Populus nigra"			
	/mol_type="mRNA"			
	/cultivar="Nxm6"			
	/db_xref="taxon:293756"			
	/clone="WS02018_N09"			
	/sex="Male"			
	/lab_host="E. coli DH10B T1 phage resistant cells"			
	/clone_lib="PTxN-IB-N-A-11"			
	/notes="Vector: pBluescript II SK (+) XR; Site 1: EcoRI (5' end of cDNA); Site 2: XhoI (3' end of cDNA); Sapling trees were exposed to continuous feeding by Cryptorhynchus lapathi (poplar and willow borer) adults caged on the sapling using mesh bags. Bark (with phloem and cambium attached) from within the caged region was harvested 2 hours, 6 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."			


```

/cultivar="Bright Yellow No.2"
/db xref="taxon:4097"
/clone="BY1636"
/cell_line="BY-2"
/clone_lib="MAT001"
/notes="Vector: pGEM-T easy; primer: M13 forward; mRNA
obtained from lag, log and stationary phase cells"

ORIGIN

Query Match          9.3%; Score 46.8; DB 5; Length 699;
Best Local Similarity 56.5%; Pred. No. 0.0087;
Matches 87; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 326 CTGAGTTGGATTCCAGCCTGTTGGTAAAGCAAAATCTTAGCCACAAAGAAATCTCTGGTC 385
DB 472 CAGTGTGGTTCAGCTTCAATCTTAAAGCAAAATCTAGCAAGAGGAGACATCTTG 413
QY 386 CGTGCAAATTCGAATAGCTTTCAATGCTCTAGAGTAGAGTGGGATTATTTATTTCG 445
DB 412 CTAGCAAGCAGCACCTTACAGTTCCGATGCACCGAGGAGGTAAACGGTGTATTCTG 353
QY 446 ATATTGCTCGATGAGAGCTGTCTCTGAGC 479
DB 352 AGGTCCGTTCTTGTGAAGAGATGTTTCATGCAGC 319

RESULT 10
CG765852
LOCUS          871 bp      DNA      linear      GSS 29-OCT-2003
DEFINITION    TcB50.3_A04 SP6 Tribolium BAC library Tribolium castaneum genomic,
genomic survey sequence.
ACCESSION     CG765852
VERSION
KEYWORDS
SOURCE        Tribolium castaneum (red flour beetle)
ORGANISM      Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 871)
Savard, J. and Tautz, D.
Tribolium castaneum BAC-ends sequencing project
Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.
Location/Qualifiers
    source          1..871
    /organism="Tribolium castaneum"
    /mol_type="genomic DNA"
    /strain="GA-2"
    /db_xref="taxon:7070"
    /clone_lib="Tribolium BAC library"
    /notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN

Query Match          8.1%; Score 40.4; DB 9; Length 871;
Best Local Similarity 52.8%; Pred. No. 0.084;
Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 167 CTCACACTGTTTAAAGCCGAAGATGCGAAACTTTTGCACAACTTTTGCATCTTTCAGAT 226
DB 644 CCCAATAGTTTTTTAGTATGAATTTTTTAAATTTTACCAATTTTGCATTTATCTGCA 703
QY 227 GCTTGACAAGAGTGTGCGAAATTTTCTCGAATATGTTTGAATAATCACCAGATCTAC 286
DB 704 ATTGCTCAAAAACACTATTAGTCGAGAACAAATATTTTTTTTGAAGAAAAAGATAATTC 763

FEATURES
    source          1..871
    /organism="Tribolium castaneum"
    /mol_type="genomic DNA"
    /strain="GA-2"
    /db_xref="taxon:7070"
    /clone_lib="Tribolium BAC library"
    /notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN

Query Match          8.1%; Score 40.4; DB 9; Length 871;
Best Local Similarity 52.8%; Pred. No. 0.084;
Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 167 CTCACACTGTTTAAAGCCGAAGATGCGAAACTTTTGCACAACTTTTGCATCTTTCAGAT 226
DB 644 CCCAATAGTTTTTTAGTATGAATTTTTTAAATTTTACCAATTTTGCATTTATCTGCA 703
QY 227 GCTTGACAAGAGTGTGCGAAATTTTCTCGAATATGTTTGAATAATCACCAGATCTAC 286
DB 704 ATTGCTCAAAAACACTATTAGTCGAGAACAAATATTTTTTTTGAAGAAAAAGATAATTC 763

```

```

QY 287 GAAGAGCCAGCCACTACTTTTCAGGCATTCACAGATAGCCCTGA 329
DB 764 AAAGAGCTTTCACAGATAGTAACCTTCATGGGNTATCTCTGA 806

RESULT 11
CG769468/c
LOCUS          846 bp      DNA      linear      GSS 29-OCT-2003
DEFINITION    TcB41.3_B05_T7 Tribolium BAC library Tribolium castaneum genomic,
genomic survey sequence.
ACCESSION     CG769468
VERSION
KEYWORDS
SOURCE        Tribolium castaneum (red flour beetle)
ORGANISM      Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 846)
Savard, J. and Tautz, D.
Tribolium castaneum BAC-ends sequencing project
Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.
Location/Qualifiers
    source          1..846
    /organism="Tribolium castaneum"
    /mol_type="genomic DNA"
    /strain="GA-2"
    /db_xref="taxon:7070"
    /clone_lib="Tribolium BAC library"
    /notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN

Query Match          8.0%; Score 40.2; DB 9; Length 846;
Best Local Similarity 56.4%; Pred. No. 0.096;
Matches 75; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 167 CTCACACTGTTTAAAGCCGAAGATGCGAAACTTTTGCACAACTTTTGCATCTTTCAGAT 226
DB 648 CCCAATAGTTTTTTAGTATGAATTTTTTAAATTTTACCAATTTTGCATTTATCTGCA 589
QY 227 GCTTGACAAGAGTGTGCGAAATTTTCTCGAATATGTTTGAATAATCACCAGATCTAC 286
DB 588 ATTGCTCAAAAACACTATTAGTCAGAACAAATATTTTTTTTGAAGAAAAATAGATTATTC 529

FEATURES
    source          1..846
    /organism="Tribolium castaneum"
    /mol_type="genomic DNA"
    /strain="GA-2"
    /db_xref="taxon:7070"
    /clone_lib="Tribolium BAC library"
    /notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN

Query Match          8.0%; Score 40.2; DB 9; Length 846;
Best Local Similarity 56.4%; Pred. No. 0.096;
Matches 75; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 167 CTCACACTGTTTAAAGCCGAAGATGCGAAACTTTTGCACAACTTTTGCATCTTTCAGAT 226
DB 648 CCCAATAGTTTTTTAGTATGAATTTTTTAAATTTTACCAATTTTGCATTTATCTGCA 589
QY 227 GCTTGACAAGAGTGTGCGAAATTTTCTCGAATATGTTTGAATAATCACCAGATCTAC 286
DB 588 ATTGCTCAAAAACACTATTAGTCAGAACAAATATTTTTTTTGAAGAAAAATAGATTATTC 529

RESULT 12
CG769307/c
LOCUS          913 bp      DNA      linear      GSS 29-OCT-2003
DEFINITION    TcB50.4_C09_T7 Tribolium BAC library Tribolium castaneum genomic,
genomic survey sequence.
ACCESSION     CG769307
VERSION
KEYWORDS
SOURCE        Tribolium castaneum (red flour beetle)
ORGANISM      Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 913)

```

AUTHORS Savard, J. and Tautz, D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
 Abteilung für Evolutionsgenetik, AG Tautz
 Institut für Genetik, Universität zu Köln
 Weyertal 121, 50931 Köln, Germany
 Tel: 49 221 470 6911
 Fax: 49 221 470 5975
 Email: savard@uni-koeln.de
 Class: BAC ends.

FEATURES Location/Qualifiers
 source
 1. .913

/organism="Tribolium castaneum"
 /mol_type="genomic DNA"
 /strain="GA-2"
 /db_xref="taxon:7070"
 /clone_lib="Tribolium BAC library"
 /note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 Library constructed by Exelixis Inc."

ORIGIN

Query Match 8.0%; Score 40.2; DB 9; Length 913;
 Best Local Similarity 56.4%; Pred. No. 0.98;
 Matches 75; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 167 CTCACACTGTTTAAAGCCGAGATGCGCAAACTTTGCCAAGTTTTCAGAT 226
 Db 476 CCCAATAGTTTTAGTTATGAATTTTAAATTTTACCAATTTTGCATTTATCTGCA 417

Qy 227 GCTTGCAACAGCATGGTGGAAATTTCTCGAATATGTTATGAAATCACCAGATCTAC 286
 Db 416 ATTTGCTCAAAACTATTAGTCAGAGAACATATTTTTTTTGAAGAAATAGATTATTC 357

Qy 287 GAAGAGCCAGCCA 299
 Db 356 AAAGAGCTTTCCA 344

RESULT 13
LOCUS CG765688/c 915 bp DNA linear GSS 29-OCT-2003
DEFINITION TC842.3_G05_SP6 Tribolium BAC library Tribolium castaneum genomic,
 genomic survey sequence.

ACCESSION CG765688

VERSION CG765688.1 GI:38018867

KEYWORDS GSS.

SOURCE Tribolium castaneum (red flour beetle)

ORGANISM Tribolium castaneum
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
 Tenebrionidae; Tribolium.

1 (bases 1 to 915)

Savard, J. and Tautz, D.

Tribolium castaneum BAC-ends sequencing project

Unpublished (2003)

Contact: Savard, J.

Abteilung für Evolutionsgenetik, AG Tautz

Institut für Genetik, Universität zu Köln

Weyertal 121, 50931 Köln, Germany

Tel: 49 221 470 6911

Fax: 49 221 470 5975

Email: savard@uni-koeln.de

Class: BAC ends.

FEATURES Location/Qualifiers

source
 1. .915

/organism="Tribolium castaneum"
 /mol_type="genomic DNA"
 /strain="GA-2"
 /db_xref="taxon:7070"
 /clone_lib="Tribolium BAC library"
 /note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 Library constructed by Exelixis Inc."

ORIGIN

Query Match 8.0%; Score 40.2; DB 9; Length 915;
 Best Local Similarity 56.4%; Pred. No. 0.98;
 Matches 75; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 167 CTCACACTGTTTAAAGCCGAGATGCGCAAACTTTGCCAAGTTTTCAGAT 226
 Db 475 CCCAATAGTTTTAGTTATGAATTTTAAATTTTACCAATTTTGCATTTATCTGCA 416

Qy 227 GCTTGCAACAGCATGGTGGAAATTTCTCGAATATGTTATGAAATCACCAGATCTAC 286
 Db 415 ATTTGCTCAAAACTATTAGTCAGAGAACATATTTTTTTTGAAGAAATAGATTATTC 356

Qy 287 GAAGAGCCAGCCA 299
 Db 355 AAAGAGCTTTCCA 343

RESULT 14

LOCUS BU357425

DEFINITION 603473713P1 CSEQCHN70 Gallus gallus cdna clone CHEST35413 5', mRNA
 sequence.

ACCESSION BU357425

VERSION BU357425.1 GI:25865426

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

1 (bases 1 to 774)

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,

Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

22335534

12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1. .774

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="White Leghorn, Hisex"

/db_xref="taxon:9031"

/clone="CHEST35413"

/dev_stage="36"

/lab_host="DH10B"

/clone_lib="CSEQCHN70"

/note="Organ: hearts; Vector: pBluescript II KS(+);

Site 1: EcoRI; Site 2: NotI; This normalized library was

constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methyated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was blunted, ligated to NotI adapters, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pBluescript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

ORIGIN

```
Query Match          7.9%; Score 39.4; DB 5; Length 774;
Best Local Similarity 59.3%; Pred. No. 1.6;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 43 TGTATCATGGATGGAAATGATCTTTTGGCTTGTATCTCTATGGGAGCCATATCTCTGG 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 586 TGGTTACTTTGATGGTCAGTATTTGGCTTGGTGGGTGTTCTCTACTAGGTTTTCTGCTG 645

QY 103 AATCCAGAGGATATGTTAGCTAGTCAAGATTGGGGAAATACCGATGAAT 155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 TTCTCTCAGAGGATTATTATTATGCAAGGTTAGGGAAGATGCCGTATAACT 698

RESULT 15
AF034173/c
LOCUS AF034173 2275 bp mRNA linear EST 30-MAR-1998
DEFINITION AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cdna
clone ntcon2 contig, mRNA sequence.
ACCESSION AF034173
VERSION AF034173.1 GI:2707735
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2275)
AUTHORS Tripodis,N. and Ragoussis,J.
TITLE Generation of a transcription map in the region immediately
centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
boundary
JOURNAL Unpublished (1997)
COMMENT Contact: Tripodis, Nikos
Division of Medical and Molecular Genetics
Guys Hospital
7th floor, Guy's Tower, London SE1 9RT, UK
Email: nikos@ki.nl.
FEATURES
source
1. .2275
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /map="6p21.3"
   /clone="ntcon2 contig"
   /clone_lib="Human mRNA (Tripodis and Ragoussis)"

ORIGIN

Query Match          7.8%; Score 39.2; DB 1; Length 2275;
Best Local Similarity 23.4%; Pred. No. 2.7;
Matches 54; Conservative 85; Mismatches 89; Indels 3; Gaps 1;

QY 188 AGATGGCAAACTTTGCCAAGTTTGCATCTTTCAGATGCTTGCACAAGCATGGTCGGA 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1701 AAAATGGCAGTGANCTCCCGAGTGACTGTTTGTGCAAAACTCYMYKAWGAAMWWWAM 1642

QY 248 AATTTTCTCGATATGTTGTAAGAAATCACCGAATCTACGAAGAGCAGCCACTACTTTT 307
      ||::: :||::: :||::: :||::: :||::: :||::: :||::: :||::: :
Db 1641 AAYTAWAAAPAAKAWYAKSWKRWAAWMAWCAWSRYRYNAKCEGCKSCWRCWRSYSS 1582

QY 308 CAGGCATTCAGATAGCCCTGAGTTGGAATTCAGCCCTTGTGGTAAAGCAAAATCTTAGC 367
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :
Db 1581 SKCRCKWY---KKTGSTSMKSAGGWRARARARARWRWYSWMSRRASARMMRYRMWGMK 1525

QY 368 CACAAGATCTCTGGTCCGTCGAATTTCTGAATTAGCTTTCAATGCTCT 418
      |::: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 1524 CYQWRRRWNGTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 1474
```

Search completed: July 12, 2005, 10:06:32
Job time : 2476.64 secs

GenCore version 5.1.6
Copyright: (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:01:48 ; Search time 2534.52 Seconds
(without alignments)
9578.171 Million cell updates/sec

Title: US-10-018-929C-2_COPY_6071_6571
Perfect score: 501
Sequence: 1 caactcatcagttactctct.....cttaaaaaaaaaaaaaaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query Match	Length	ID	Description		
1	501	100.0	6571	6	AX063583		
2	489.4	97.7	10329	6	AX063582		
3	489.4	97.7	10329	8	AF213628		
4	489.4	97.7	120965	8	AC026875		
5	484	96.6	6554	8	AF213627		
6	443	88.4	939	8	AK117959		
7	245	48.9	702	8	BT006219		
8	59.8	11.9	574	6	AX063612		
9	51.6	10.3	227751	2	CR538727		
10	51.2	10.2	340552	3	PFA929354		
11	50.8	10.1	162905	2	BX927388		
12	50.6	10.1	201851	2	CR743181		
13	50.6	10.1	242492	2	BX950852		
14	50.6	10.1	264994	2	CR628325		
15	50.6	10.1	277724	2	CR450772		
16	48.8	9.7	164399	3	PFMA13P6		
17	47.8	9.5	163724	2	CR788299		
18	46.6	9.3	3616	3	BT003763		
19	46.4	9.3	124330	3	PFMA11P1		

20	46	9.2	164335	2	BX950865	BX950865	Danio rer
21	45.8	9.1	254050	3	PFA929358	AL929358	Plasmodiu
22	45.4	9.1	347050	3	PFA929351	AL929351	Plasmodiu
23	45	9.0	169504	2	CR381556	CR381556	Danio rer
24	45	9.0	191689	2	BX957320	BX957320	Danio rer
25	44.8	8.9	110000	2	PFMA18P1_02	Continuation (3 of	
26	44.8	8.9	159003	2	BX571978	BX571978	Danio rer
27	44.6	8.9	33651	3	AC114261	AC114261	Dictyoste
28	44.6	8.9	38881	2	AC149371	AC149371	Phakopor
29	44.6	8.9	212955	2	AC136676	AC136676	Rattus no
30	44.6	8.9	253151	3	AE014842	AE014842	Plasmodiu
31	44.4	8.9	20063	8	CGL511533	AJ511533	Candida g
32	44.4	8.9	58454	5	BX784394	BX784394	Zebrafish
33	44.4	8.9	61052	2	AC123513	AC123513	Dictyoste
34	44.4	8.9	136240	3	AC117070	AC117070	Dictyoste
35	44.4	8.9	163443	2	AC066280	AC066280	Plasmodiu
36	44.4	8.9	253132	3	AE014846	AE014846	Plasmodiu
37	44.2	8.8	214310	2	CR847982	CR847982	Danio rer
38	44.2	8.8	266544	3	AC116956	AC116956	Dictyoste
39	44	8.8	110158	9	AC092663	AC092663	Homo sapi
40	44	8.8	166239	2	AC078965	AC078965	Homo sapi
41	44	8.8	226450	2	CR522882	CR522882	Danio rer
42	44	8.8	244097	2	CR354394	CR354394	Danio rer
43	43.8	8.7	110000	3	AC116957_1	Continuation (2 of	
44	43.8	8.7	154071	3	AC115598	AC115598	Dictyoste
45	43.8	8.7	333321	3	AC116986	AC116986	Dictyoste

ALIGNMENTS

RESULT 1
AX063583
LOCUS AX063583 6571 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 2 from Patent WO0100801.
ACCESSION AX063583
VERSION AX063583.1 GI:12541307
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses.
REFERENCE 1
Habu,Y., Mittelsten Scheid,O., Amedeo,F. and Paszkowski,J.
Gene involved in epigenetic gene silencing
Patent: WO 0100801-A 2 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

FEATURES
source
1. .6571
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
310. .6315
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC25098.1"
/db_xref="GI:12541308"
/translation="MKKDEKIGLTGRTIYTRSLAASIPASVEQETPLGRSSRCPTST
KVITPASATRSRLAPSPASVSKSGGIVKNSPTPSLRKSNRKTQSVLSQSKGSDN
SVRGKGTSPDIEQRKDSVEESTDKIRPIMSARSYRALFRGLKESALVSNPSEEL
VIVGCGRRIPAGNDVQGTCDPPADAGSKRLPVDTSLDKGTDFPLKSVTEKEIV
LDASPIVETGDSVIGSPENLETKLQDKTDCSPANAESKTLPVGETSLEKEYPQ
KFDNDNTDCLPPANASKRLPVGETSLEKDTDFPLKSTTGKWLVSPIVETRDS
VISCPSNTLETQKLVSKTGLETDIVLPLKRGDTABIEDACATVANGDDHVNSSDG
VTSREYENGQIQHKSDEPKFSYVPEYVQVLSQLEQVQCTLFPSKSLSSL
SKIDLAGLETLNSVRKTDHPYVMDASLKQLLTKNLEHLDEIVKASGKLHLDK
MLTHIKNGLKAVVFOATOTPEGLIGNLEDFVGORFGPKSVHEGIVYSSKNAIN
NPNKESQCCVILLETACSTIKLRADAFILFGSSINPSHDVHKVKKIESCSERT
KIFRLSVCTVEERKALILARQNQRKAVENLRSLTHALLMGASVYDFDKLDHFHS
ETPDGVSFQESIMDGVTHIEFSSILSSKGEENEVKLCLLLEAKHQAGTYSDDSTLFG

EDHILKLSDBSPNIFWSKLLGGKPNMVKPSDTPORNRKRVQYFEGSEASPKTGDCGN
AKRKAAGDDVDPRTVDDPDDDERKASGDHMGALSPKVTITLQSSCKSSGCTGL
DGNDAFLGSMGSHIGIEDMLASDQWGIPIESORRLHTVLKPMKAKCQVHLHSD
ACTSMVGNFVYVIEHRIYEEPAITDQFQIALSWIAALLVKQILSHAKESIVRANSE
LAPKCRSEVDYIYSLSCWKSFLBHTQLOPFCFTNSKQSVSTKLIVNESLSGAT
VRDEKINTKSMRNSDEECMTKRSCHYSTATRDIEKTIISGKKYKQVOKLQVGH
EEKMELLAMYADKKQKLETSKVEAAVIRITCSRTSTOVGDILKLHDHYVERKFDIK
SEKVECKSLSEOMHEVAKKLADEACWLNRIKSWAAKLKVCVPIQSGNNKHFSGSSN
ISQNAEDVOICNNANVEATYADNCHASKVNPQVPEANTILGWSGSGTOQVHEMVDVR
NDETMDVLSALSRQLTKSQSNHASITFPEILIPADQCBFAALNVHLBEDQNCDRIT
SAASDSDSRVPEVQSLENHASITFPEILIPADQCBFAALNVHLBEDQNCDRIT
NREDCLDIOBIPDELAMPVQHVLASVVEVTEFAASHQGDVTCPLSSPTGNQPAPEANT
TESNLEEAIEPOSAGSETVETDFAASHQGDVTCPLSSPTGNQPAPEANT
NTSAEPHAGDAVESGDVAIDQETMGADQACSLPSGSVGTQSDIGANIEGONVTV
AQLPDGSADVTVGGSFVSDQACDASPMPLSPGNHPTDPAVNIENGLDNTSVAPRHS
GLDACMEISEPQOVERSTFANLFHEGGVAGVATVPLSLNNGNTQIOAVPVPQ
IPEPNDPFLHEKLEKRESENKATPEEKSKILKAELEKMAEQVAFRKRHFVE
AENHTTKIEKNDKIVIMNKLJANAFSLKCTDKVSPGAPRGTQQLAQRAAQVSA
LRNYIAPOOLOASFPAPALVSPALQLOQSSFPAPGPAPLOQASSFPSSVSRPSALL
LNFVCPMPQBPQPLISNIAPTSPVTPATNPGLSRPAHLNSYRPPSSSTPVATPTS
SVPPQALTVSAVSIQQQEQPOQSLSSGLOSNNVVCUSDDE"

ORIGIN

Query Match		100.0%;	Score 501;	DB 6;	Length 6571;
Best Local Similarity		100.0%;	Pred. No. 1.2e-115;		
Matches 501;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CAACTCCATCAGTTACTCTCGCAACAAATCCAGGTCTGCGTTCTCTGCAACCACTAA	60		
DB	6071	CAACTCCATCAGTTACTCTCGCAACAAATCCAGGTCTGCGTTCTCTGCAACCACTAA	6130		
QY	61	ACTCATATAGACCATCTCTTCACTCCCGTCGCCACAGCTACTCCAACTCTGTCAGTGC	120		
DB	6131	ACTCATATAGACCATCTCTTCACTCCCGTCGCCACAGCTACTCCAACTCTGTCAGTGC	6190		
QY	121	CTCCTCAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACAACCGC	180		
DB	6191	CTCCTCAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACAACCGC	6250		
QY	181	ACAGAGCTTGAGCAGTGGATTCGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG	240		
DB	6251	ACAGAGCTTGAGCAGTGGATTCGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG	6310		
QY	241	AGTGACCTTAAGAGAGAGATGGTTAGGCTCTAGTTATTCATTTTAGAGAGTTAATAAT	300		
DB	6311	AGTGACCTTAAGAGAGAGATGGTTAGGCTCTAGTTATTCATTTTAGAGAGTTAATAAT	6370		
QY	301	AGTATATATATATATATAGTAGGTACCTAATCTCTGCTGTTAATCTAATTTTAGTGA	360		
DB	6371	AGTATATATATATATATAGTAGGTACCTAATCTCTGCTGTTAATCTAATTTTAGTGA	6430		
QY	361	GTGAGAACCGACTCGTTGGCTGAAGTCTCTCTTTTGAACCGAAGCTTCTACTTTTCAT	420		
DB	6431	GTGAGAACCGACTCGTTGGCTGAAGTCTCTCTTTTGAACCGAAGCTTCTACTTTTCAT	6490		
QY	421	GTATATAATAACAGTCTGATCACACACAAATTTGATGATTCGAAATACTACTGATTTA	480		
DB	6491	GTATATAATAACAGTCTGATCACACACAAATTTGATGATTCGAAATACTACTGATTTA	6550		
QY	481	ACTTAAAAAATAAAAAAAAAAAAAA	501		
DB	6551	ACTTAAAAAATAAAAAAAAAAAAAA	6571		

RESULT 2

AX063582
LOCUS AX063582 10329 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100801.
ACCESSION AX063582
VERSION AX063582.1 GI:12541306
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustersids II; Brassicales; Brassicaceae; Arabidopsiis.
1
Habu, Y., mittelsten Scheid, O., Amedeo, P. and Paszkowski, J.
Gene involved in epigenetic gene silencing
Patent: WO 0100801-A 1 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)

FEATURES		Location/Qualifiers
source		1..10329 /organism="Arabidopsis thaliana" /mol_type="unassigned DNA" /db_xref="taxon:3702"
intron		1009..1295
intron		2551..2673
intron		2753..2867
intron		3114..3506
intron		3681..3973
intron		4896..4975
intron		5218..5777
intron		5883..6082
intron		7481..7615
intron		7772..7914
intron		8071..8153
intron		8319..8451
intron		8630..8718
intron		8919..9000
intron		9212..9284
ORIGIN		

Query Match		97.7%;	Score 489,4;	DB 6;	Length 10329;
Best Local Similarity		98.8%;	Pred. No. 1e-112;		
Matches 493;		Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	CAACTCCATCAGTTACTCTCGCAACAAATCCAGGTCTGCGTTCTCTGCAACCACTAA	60		
DB	9564	CAACTCCATCAGTTACTCTCGCAACAAATCCAGGTCTGCGTTCTCTGCAACCACTAA	9623		
QY	61	ACTCATATAGACCATCTCTTCACTCCCGTCGCCACAGCTACTCCAACTCTGTCAGTGC	120		
DB	9624	ACTCATATAGACCATCTCTTCACTCCCGTCGCCACAGCTACTCCAACTCTGTCAGTGC	9683		
QY	121	CTCCTCAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACAACCGC	180		
DB	9684	CTCCTCAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACAACCGC	9743		
QY	181	AACAGAGCTTGAGCAGTGGATTCGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG	240		
DB	9744	AACAGAGCTTGAGCAGTGGATTCGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG	9803		
QY	241	AGTGACCTTAAGAGAGAGATGGTTAGGCTCTAGTTATTCATTTTAGAGAGTTAATAAT	300		
DB	9804	AGTGACCTTAAGAGAGAGATGGTTAGGCTCTAGTTATTCATTTTAGAGAGTTAATAAT	9863		
QY	301	AGTATATATATATATATAGTAGGTACCTAATCTCTGCTGTTAATCTAATTTTAGTGA	360		
DB	9864	AGTATATATATATATATAGTAGGTACCTAATCTCTGCTGTTAATCTAATTTTAGTGA	9923		
QY	361	GTGAGAACCGACTCGTTGGCTGAAGTCTCTCTTTTGAACCGAAGCTTCTACTTTTCAT	420		
DB	9924	GTGAGAACCGACTCGTTGGCTGAAGTCTCTCTTTTGAACCGAAGCTTCTACTTTTCAT	9983		
QY	421	GTATATAATAACAGTCTGATCACACAAATTTGATGATTCGAAATACTACTGATTTA	480		
DB	9984	GTATATAATAACAGTCTGATCACACAAATTTGATGATTCGAAATACTACTGATTTA	10043		
QY	481	ACTTAAAAAATAAAAAAAAAAAAAA	499		
DB	10044	ACTTAAAAAATAACCCAAA	10062		

RESULT 3

NLWEDLGKAMEGDEIEVGEVDCGTSTRAVCTKVEIHSYPTFMIFYNGEESVSKYKGRD
VESUKAFVVEIEKAAKAQLEDKEL"
complement (join(25423, .25563,25675, .26048,26127, .26330,
26585, .26783,26941, .27842,27936, .27991,28208, .28413))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T6D22.6"
/protein_id="AAF79845.1"
/db_xref="GI:8778846"
/translacion="MDSVRRKINNTTQSSPSKPSKFSVYRNPALAAASTANSIRP
SKSVLLFIILLSIASAFSLISFMAGEKMLTNAITFGKISOEAAYVPAKQGLVTLPC
IGAMWLSKGISLHRAKPAKGETSKFSTKQOFSLSNRQLELLGKKADQSVSEF
PSKYPALQSLPVLPIHQTUTGSAHSSGGDKLNSRNGSQISFSTPSKQMS
KSVLLPSSSPVSNLUSSGQDRAVSPFMSGRSSAKDITTEQLEQLLAIIDEKITE
SAGMRTPTPTVGFAMAFSTVGGSCAGATSTPLRPMSPGCAKFTTPTPKGE
GDFNPMSLRATBEGFTHLGVYPOIEDWRDLROWSCSVLLKPLNKVETSHLOVMOT
ASKLGVNVTVOGSDLPNTGTATTALPVDRTKGWPYSILDEDAHLAHLANLVOAI
DASMQKLRTENQFQQQQQQQQAALI PVQECVDALSHRRLGLMKGEVWGLLP
RSSLPADYTVQRIKALAEETCVKNYENGADTKRKNWSLEPPTDLSHLLLYLFCAP
LEHPKMWHLDPSSYCTQASKNPLFLGLVLPKPEKPEKVIYVGVSPSTLHFGACVL
AVDKQSPPTFALYWDKKVQFTLOQRTALWDSMLLI CHIRIKVGVGVVVRGNLGSALN
ILOVDSDDTD"
complement (join(29418, .29491,29579, .29620,29858, .29942,
30386, .30805))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T6D22.7"
/protein_id="AAF79844.1"
/db_xref="GI:8778845"
/translacion="MWSKKPKKARSDVVVNKASGRSKRSGSRFKTKSNKNIYVK
KPEIYBISSSSDSYEEAIRGDGAKNKSVVSKNGKSGVGIPTKTSKNREDDGG
AEDAKIFPMNRRIRIRSDNSAPQIMQDAVFLVKNKATEMFIERFSEAYDSSVKDKK
KFTHYKHLSSVNSDQRYEFLADSVPEKLKAAALBEWEGMTDAG"
Query Match 97.7%; Score 489.4; DB 8; Length 120965;
Best Local Similarity 98.8%; Pred. No. 1.1e-112;
Matches 493; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CAATCCATCAGTTACTCTCGCAAAATCAGGTCTGCGTTCTCTCGCACCACTAA 60
DB 58120 CAATCCATCAGTTACTCTCGCAAAATCAGGTCTGCGTTCTCTCGCACCACTAA 58061
QY 61 ACTCATATAGACCATCTCTTCACTCCGTCGACAGTACTTCAACCTCGTCAGTGC 120
DB 58060 ACTCATATAGACCATCTCTTCACTCCGTCGACAGTACTTCAACCTCGTCAGTGC 58001
QY 121 CTCTCAAGCTTTGACATATTCAGTGTGCAATTCAGCAGCAGCAGCAACAACCGC 180
DB 58000 CTCTCAAGCTTTGACATATTCAGTGTGCAATTCAGCAGCAGCAGCAACAACCGC 57941
QY 181 AACAGACCTTGAGCAGTGGATTGACAGACCAATGAAGTGGTTTGTCTTCTGACGACG 240
DB 57940 AACAGACCTTGAGCAGTGGATTGACAGACCAATGAAGTGGTTTGTCTTCTGACGACG 57881
QY 241 AGTGACCTTAAGAGAGAGATGGTTAGGTCTTGTAGTTATGATTTTATAGAGATTAAAT 300
DB 57880 AGTGACCTTAAGAGAGAGATGGTTAGGTCTTGTAGTTATGATTTTATAGAGATTAAAT 57821
QY 301 AGTATATATATATATATAGTAGTAGTTACTTAATCTCTGTCGTTAATCTAATTTAGTGA 360
DB 57820 AGTATATATATATATATAGTAGTAGTTACTTAATCTCTGTCGTTAATCTAATTTAGTGA 57761
QY 361 GTCAGGACCGACCTCGTTGGTGAAGTCTCTCTTTTGAACGCAAGCTTCTACTTTTCAT 420
DB 57760 GTCAGGACCGACCTCGTTGGTGAAGTCTCTCTTTTGAACGCAAGCTTCTACTTTTCAT 57701
QY 421 GTATATAAATACAGTCTGATCACACAACAAATTTGATGTTTGAATACTACTGATTTA 480
DB 57700 GTATATAAATACAGTCTGATCACACAACAAATTTGATGTTTGAATACTACTGATTTA 57641
QY 481 ACTTAAAAAIAAAAAAAAAA 499

||||| ||||| ||||| ||||| |||||
Db 57640 ACTTATAGAAACCCAAA 57622
RESULT 5
AF213627
LOCUS Arabidopsis thaliana MOM mRNA linear PLN 01-JUN-2000
DEFINITION Arabidopsis thaliana MOM mRNA, complete cds.
ACCESSION AF213627
VERSION AF213627.1 GI:8132767
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsais.
1 (bases 1 to 6554)
Amedeo, P., Habu, Y., Afsar, K., Scheid, O.M. and Paszkowski, J.
Disruption of the plant gene MOM releases transcriptional silencing
of methylated genes
NATURE 405 (6783), 203-206 (2000)
JOURNAL MEDLINE 20279299
PUBMED 10821279
REFERENCE 2 (bases 1 to 6554)
Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. and
Paszkowski, J.
Direct Submision
Submitted (09-DEC-1999) Friedrich Miescher Institute, P.O. Box
2543, Basel 4002, Switzerland
FEATURES
Location/Qualifiers
source
1. 6554
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Zhuerich"
/db_xref="taxon:3702"
/chromosome="1"
/map="between ATEAT1 and phyA"
310..6315
/function="involved in the maintenance of transcriptional
gene silencing"
/codon_start=1
/product="MOM"
/db_xref="GI:8132768"
/db_xref="AAF73380.1"
CDS

JOURNAL Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:mseki@sec.riken.go.jp) URL:http://pfwebweb.gsc.riken.go.jp, Tel:81-45-503-9625, Fax:81-45-503-9586) COMMENT An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda ELC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://pfwebweb.gsc.riken.go.jp/) for further details. FEATURES source 1. .939 /organism="Arabidopsis thaliana" /mol_type="mRNA" /db_xref="taxon:3702" /chromosome="1" /clone="RAFL19-14-009" /ecotype="Columbia" /notes="common name: thale cress" gene 1. .939 /genes="At1g08060/T6D22_27" CDS 40. .741 /genes="At1g08060/T6D22_27" /codon_start=1 /product="putative storage protein" /protein_id="BAC42596.1" /db_xref="GI:26450979" /translation="MAEVAQAEFRKFEVEAEHNTTRTKIEKDNLVIMKLANAPL SKTKDKVSPSGAPRGKI QOLQORAQVSAALRYIAPOOLQASFPALVAPLOLQ QSFPPAGPAPLOQASFPSSVSRPSALLNFAVCPMPQPROPLISNTAPSVTPA THPLRSAPHLNYSRPSSTPVATPTSSVPPQALTVSAVSIQQOQQEQPQSSLS GLQSNNEVVCLSDDE" ORIGIN Query Match 88.4%; Score 443; DB 8; Length 939; Best Local Similarity 100.0%; Pred. No. 4.8e-101; Mismatches 0; Gaps 0; Matches 443; Conservative 0; Indels 0; Gaps 0; Qy 1 CAATCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGGGTTCTCTGACACCACTAA 60 Db 497 CAATCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGGGTTCTCTGACACCACTAA 556 Qy 61 ACTCATATAGACCATCTCTTCAACTCCCGTCCGACAGCTACTCCAACTCTGTCAGTGC 120 Db 557 ACTCATATAGACCATCTCTTCAACTCCCGTCCGACAGCTACTCCAACTCTGTCAGTGC 616 Qy 121 CTCTCAAGCTTTGACATATTAGCTGTGTCAATTCAGCAGCAGCAAGAACCAACCCG 180 Db 617 CTCTCAAGCTTTGACATATTAGCTGTGTCAATTCAGCAGCAGCAAGAACCAACCCG 676 Qy 181 AACAGAGCTTGACAGTGGATTGCAGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG 240 Db 677 AACAGAGCTTGACAGTGGATTGCAGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG 736 Qy 241 AGTGACCTAAGAGAGAGATGGTTAGGCTTTAGTTATTGATTTTATAGAGAGTTAAT 300 Db 737 AGTGACCTAAGAGAGAGATGGTTAGGCTTTAGTTATTGATTTTATAGAGAGTTAAT 796 Qy 301 AGTATATATATATATATATAGTATAGTTAGTTACTTAATCTCTCGTTAATCTAATTTAGTGA 360 Db 797 AGTATATATATATATATATAGTATAGTTAGTTACTTAATCTCTCGTTAATCTAATTTAGTGA 856 Qy 361 GTCAGGAACCGACTCGTTGGCTTAAGGTCTCTCTTTTGAACCGAAGCTTCTACTTTTAT 420 Db 857 GTCAGGAACCGACTCGTTGGCTTAAGGTCTCTCTTTTGAACCGAAGCTTCTACTTTTAT 916 Qy 421 GTATATAAATACAGTCTGATCAC 443 Db 917 GTATATAAATACAGTCTGATCAC 939

JOURNAL Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:mseki@sec.riken.go.jp) URL:http://pfwebweb.gsc.riken.go.jp, Tel:81-45-503-9625, Fax:81-45-503-9586) COMMENT An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda ELC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://pfwebweb.gsc.riken.go.jp/) for further details. FEATURES source 1. .939 /organism="Arabidopsis thaliana" /mol_type="mRNA" /db_xref="taxon:3702" /chromosome="1" /clone="RAFL19-14-009" /ecotype="Columbia" /notes="common name: thale cress" gene 1. .939 /genes="At1g08060/T6D22_27" CDS 40. .741 /genes="At1g08060/T6D22_27" /codon_start=1 /product="putative storage protein" /protein_id="BAC42596.1" /db_xref="GI:26450979" /translation="MAEVAQAEFRKFEVEAEHNTTRTKIEKDNLVIMKLANAPL SKTKDKVSPSGAPRGKI QOLQORAQVSAALRYIAPOOLQASFPALVAPLOLQ QSFPPAGPAPLOQASFPSSVSRPSALLNFAVCPMPQPROPLISNTAPSVTPA THPLRSAPHLNYSRPSSTPVATPTSSVPPQALTVSAVSIQQOQQEQPQSSLS GLQSNNEVVCLSDDE" ORIGIN Query Match 88.4%; Score 443; DB 8; Length 939; Best Local Similarity 100.0%; Pred. No. 4.8e-101; Mismatches 0; Gaps 0; Matches 443; Conservative 0; Indels 0; Gaps 0; Qy 1 CAATCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGGGTTCTCTGACACCACTAA 60 Db 497 CAATCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGGGTTCTCTGACACCACTAA 556 Qy 61 ACTCATATAGACCATCTCTTCAACTCCCGTCCGACAGCTACTCCAACTCTGTCAGTGC 120 Db 557 ACTCATATAGACCATCTCTTCAACTCCCGTCCGACAGCTACTCCAACTCTGTCAGTGC 616 Qy 121 CTCTCAAGCTTTGACATATTAGCTGTGTCAATTCAGCAGCAGCAAGAACCAACCCG 180 Db 617 CTCTCAAGCTTTGACATATTAGCTGTGTCAATTCAGCAGCAGCAAGAACCAACCCG 676 Qy 181 AACAGAGCTTGACAGTGGATTGCAGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG 240 Db 677 AACAGAGCTTGACAGTGGATTGCAGAGCAACAAATGAAGTGGTTGTCTTTCTGACGACG 736 Qy 241 AGTGACCTAAGAGAGAGATGGTTAGGCTTTAGTTATTGATTTTATAGAGAGTTAAT 300 Db 737 AGTGACCTAAGAGAGAGATGGTTAGGCTTTAGTTATTGATTTTATAGAGAGTTAAT 796 Qy 301 AGTATATATATATATATATAGTATAGTTAGTTACTTAATCTCTCGTTAATCTAATTTAGTGA 360 Db 797 AGTATATATATATATATATAGTATAGTTAGTTACTTAATCTCTCGTTAATCTAATTTAGTGA 856 Qy 361 GTCAGGAACCGACTCGTTGGCTTAAGGTCTCTCTTTTGAACCGAAGCTTCTACTTTTAT 420 Db 857 GTCAGGAACCGACTCGTTGGCTTAAGGTCTCTCTTTTGAACCGAAGCTTCTACTTTTAT 916 Qy 421 GTATATAAATACAGTCTGATCAC 443 Db 917 GTATATAAATACAGTCTGATCAC 939


```
/gene="PF01200w"
join(1232, .4857, 4998, .5103)
/gene="PF01200w"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51604.1"
/db_xref="GI:23504823"
/db_xref="UniProt/TREMBL:Q813M4"
/translation="MTRKRVKGLDNNNIISSYINNDVNNKKNSYKNEENPYSKRRK
KINDEHRNSTCDVNNKYFTDENISDEQNTCDNSELSDNDDDVYVG
RSDASSYKKRKKKIIKKRNNNNKTBINKVKEKEKQSSNENFPKKDER
KQPKKKKQNTILEKWHNFIYOKNEVFDLDFKYLSTLTERDEEKNEHFSFVLS
EFLNMLRCKKKNQGEVEKQINDDNIHIDNIIHIDNIIHIDNIIHIDNIIHID
HIDCONVKNNSSENIKKERESDISCEVSNFISKIFNKENIDLCILNDN
TNMNIFFLEDIQIHLNMLTKVPSFSSRVVDNIKKSYLLKKKKNNFHKQKVEQ
NHDFNYLEKIYINGOKNRQSEIREERRGNNDFOECNINIEGNNILHDPDIVF
LYFLICNSLNDIKENTSEENSIKTNINEETLNIENLIDLEVLKSLDSSENI
PLSLCCPNYMFDSINMVTYLPFLNEKNMDKLYNKDTPYINISYNNINNSI
SGEVLDDKRLNENYNNNNNSISGEVLYDDKNTLKNNNYRSIINKDVTNP
RYENDNIYKNIENIYNLNDMMKCEIPEGNILSIYSIYLSNDAYYNSDKPGR
NFIVFYSKYLCPVFDIITIEHTYMKCKMEWIKPDSLAKGKRGKEIYILLCLL
GNKIKIPASIKYHFFVNIYMALKNQNRKDSNNIYSFINKERKKEKKQKKTQKROK
FKIPTVTSYINDEFYTCIKDKQRYLKTALCFNNFKIIDIILNIVNKEENKIN
SFIEKSTINISLOYGREHYENFLFNKQKNINQNMQSEGNINENDELVHRSWN
KEMKKENFFPSSWENKNDVNNHDDAINCHNVDCINDRNNNLIISDKRKFLETH
NISRGKINNYINISICLNINKINIRIEIYLOHGVLSLCSFYPCINSYLLCLCS
KDNIVIDIRNSKELFYFKRTCTSHLKYWRNSCISFGQDKGSIINLFQNKYILNI
DKSPNTVNDLSVIGDMHLYFLYDDNTILKGTKNSSKNKYNIEFFLWSTKQVD
LEPSILLESQKQEDIVSYMILLYVGLQNLVNDGKIIVSKKWDQOCTGKNA
LTPKSIYASDENIYATGSCGLIHLFSGEKDENNV"
complement(join(5631, .5794, 5949, .6017, 6168, .6399))
/gene="PF01205c"
complement(join(5631, .5794, 5949, .6017, 6168, .6399))
/gene="PF01205c"
/note="4 probable transmembrane helices predicted by
TMHMM2.0 at aa 7-26, 46-65, 78-97 and 112-129"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51605.1"
/db_xref="GI:23504824"
/db_xref="UniProt/TREMBL:Q813M3"
/translation="MKSCFVNKFVASGVGVILVLYEFSQYENIYKYNHVVYIMDN
SYIPFTFLSLSYMLNIFSPKSNISNSFGLLGCLITCSMFYFMQVITCSY
AKESTKFLRSVYCVFLYFMSLCTESYKNFSFTTSKYERKEPGTMKEV"
complement(7946, .9166)
/gene="PF01210c"
complement(7946, .9166)
/gene="PF01210c"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51606.1"
/db_xref="GI:23504825"
/db_xref="UniProt/TREMBL:Q813M2"
/translation="MHEELHIVLSNEKDVENFIVGLNKKFVLEDYNSDKYKIT
YNDKREKNEDEEGNVIQNSIKIIVNNKYISADVISSCIYNKILNVEKSHIDNI
PYESIIFLDDFNKRNKIIINHPFNRYNNDQCMYIKDKNGMIRVNFDFSNLYI
DIGIKAIPLSLHEKNTDILYFSEYFIALYINYNLIEQEIPEDEQDNDNKN
NENKNNKNIINKKPEEYEDDERLVEALCNMWKGLQMKKANIIMHLENNLON
QDKKEEDTKNDDORADENKTKQNKEDYQNVLTQDTELDKNYIEHKGEQTYCQIQ
KEEQNKCIQKEEQKQVQKEDLFMQNFNIVEKIQVKIENTCNDNVRKKAAE
LILELQYFCDDMBE"
complement(10448, .11545)
/gene="PF01215c"
complement(10448, .11545)
/gene="PF01215c"
/note="Similar to Homo sapiens developmentally regulated
GTP-binding protein 1 drgl SW:DRGI_HUMAN (Q9Y295) (367 aa)
fasta scores: E(): 8.6e-100, 71.858% id in 366 aa, and to
Arabidopsis thaliana GTP-binding-like protein f23k16.150
or at4g39520 TR:Q9SVA6 (EMBL:AL078620) (369 aa) fasta
scores: E(): 1.3e-99, 69.945% id in 366 aa"
/codon_start=1
/product="developmentally regulated GTP-binding protein 1,
```

```
putative"
/protein_id="CAD51607.1"
/db_xref="GI:23504826"
/db_xref="UniProt/TREMBL:Q813M1"
/translation="MSILOKTIADIEAEMAKTKATNYHLGLLKAKLSKLAOLIEG
GTKGGEGEGFDVSKTGDARIGLVGFPVSGKSTLLNKLGTFSVASVEPTLTCVPG
IFKYGANQQLLDLPFIIEGAKDGRGKQVIAVAKSCSLILVLDVLKPUTYRKIIIE
KEBEGFGRINLKKPFIIFQKKDGGINITHVPLNLDDEMIKISICHEYRIINANIS
IRCBATVDDIIVIEGRLNVPCYIYVNLKVDQITMEELNVLTKLPHNVPIGAHLENNL
DGLLEAIVNTYLDLVRITVTKPGQIPDYESPVLKKKCKVENFCKIHRSLVQQLKVA
LVMGSKVKHNPQKVGKDHENDEVDVQLVK"
13193, .13564
/gene="PF01220w"
13193, .13564
/gene="PF01220w"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51608.1"
/db_xref="GI:23504827"
/translation="MEGLYDKYDLRNNNISLLEMSSEEDLKVINNNKNLVEYYNNI
LNKQITVNTYFNMDITDDIVEMNNAFSDPLKINEESDDNVEYKINGFLSLYNKPSS
GTKVLYDHMDITVHCEQEQN"
14804, .15577
/gene="PF01225w"
14804, .15577
/gene="PF01225w"
/note="Similar to Porphyra purpurea chloroplast 50s
ribosomal protein l12 rpl12 SM:RK12 PORPU (P51339) (129
aa) fasta scores: E(): 1.1e-09, 42.222% id in 135 aa
Signal anchor predicted by SignalP 2.0 HMM (Signal peptide
probability 0.000, signal anchor probability 0.837) with
cleavage site probability 0.000 between residues 60 and 61
1 probable transmembrane helix predicted by TMHMM2.0 at aa
35-57"
/codon_start=1
/product="50S ribosomal subunit protein l12, putative"
/protein_id="CAD51609.1"
/db_xref="GI:23504828"
/db_xref="GOA:Q813L9"
/db_xref="UniProt/TREMBL:Q813L9"
/translation="MKSIVGSKVTKLICKIKCKEKNRRKNNKTKDCSLFFIILKQKS
CPLLYILSLIYITMECIVCMKLSKNDNCLNNKFFYLEIKLNNDIYGGKSVNNI
RHKISQDNNVNNHNFJKSSKVDEIESLKNLTLLESELVKKILTLTFSVDTKQNL
NANSQCKENQKSDGKNEEENKVDILILENIEPNKKIPIIKIVKEIKDLNKLK
QAKDLVDNLPQTLFEKVNETADKTKLTAAGGIVKLK"
complement(16580, .16897)
/gene="PF01230c"
complement(16580, .16897)
/gene="PF01230c"
/note="Similar to Arabidopsis thaliana mitochondrial
import receptor subunit tom22 at5g43970 TR:Q9FNC9
(EMBL:AB006703) (99 aa) fasta scores: E(): 0.29, 34.146%
id in 41 aa
Putative conserved hypothetical protein
1 probable transmembrane helix predicted by TMHMM2.0 at aa
58-77"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="CAD51610.1"
/db_xref="GI:23504829"
/db_xref="UniProt/TREMBL:Q813L8"
/translation="MGTALSKIITINEENRLVISNKPVFOLAHKDAIMRNKLRVAKN
RINFLKKSIKTTSVVVVIAGVSVVLTPIAFQYKECQLFENQAQFQAQAAVNP
QLN"
complement(join(18403, .20244, 21042, .21302))
/gene="PF01235c"
complement(join(18403, .20244, 21042, .21302))
/gene="PF01235c"
/note="Weakly similar to Arabidopsis thaliana similar to
ufdi protein atg15420 TR:Q23395 (EMBL:Z97338) (778 aa)
fasta scores: E(): 1.3e-08, 20.786% id in 611 aa
Hypothetical protein"
```



```
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51611.1"
/db_xref="GI:23504830"
/db_xref="GOA:Q813L7"
/db_xref="UniProt/TREMBL:Q813L7"
/translations="MDDDFVRLNKNKSFSPKKNKSLNEVLQKNKRNDDKEDINCN
KLKNDLEQLQVHSSNNKICQKFLTLPLNKKSDKLKDHSDKIILPVSLKTLKGNVKN

Query Match      10.2%; Score 51.2; DB 3; Length 340552;
Best Local Similarity 51.9%; Pred. No. 0.041;
Matches 139; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 235  ACCGACGAGTCAACGAGAGAGAGAGTGGTATAGGCTCTTAGTATTGATGATTTTAGAGAGTT 294
DB 195810  ATGAAAATGAAATGAAAATGAAATATATATACAAATATACATAGTATGATTCATAT 195751

QY 295  AATAATAGTATATATATATATATATATATATATATATATATATATATATATATATATAT 354
DB 195750  TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195691

QY 355  TAGTGAG-TCAGGACCGACTCGTTGGCTAAGGTCTCTCTTTTGAACGCAAGGTCTTA 413
DB 195690  TTTTCAGTTTATAATTTTTTTCTTATACACTATATCTGATTTTATAAATAAATATATA 195631

QY 414  CTTTCATGTATATAATACAGTCTGTATCACACACAAATTTGATGATTCGAAATACTAC 473
DB 195630  AATATATTAATATATATATATATATATATATATATATATATATATATATATATATATAT 195571

QY 474  TGATTTAACTTAAAAAAAATATATATATATATATATATATATATATATATATATATATAT 501
DB 195570  TCATTTAAAAAATATATATATATATATATATATATATATATATATATATATATATATAT 195543

RESULT 11
BX927388      162905 bp      DNA      linear      HTG 04-FEB-2004
LOCUS      Danio rerio clone CH211-160J14, *** SEQUENCING IN PROGRESS ***, 9
DEFINITION      unordered pieces.
ACCESSION      BX927388
VERSION      BX927388.3 GI:42406633
KEYWORDS      HTG; HTGS PHASE1.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
              McRay, K.
              Direct Submission
              Submitted (03-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On Feb 4, 2004 this sequence version replaced gi:41630145.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              ----- Project Information
              Center project name: zC160J14
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 160829 bases at least Q40
              Consensus quality: 161281 bases at least Q30
              Consensus quality: 161664 bases at least Q20
              Insert size: 162105; sum-of-contigs
              Insert size: 167052; 4.8% error; agarose-1p
              Quality coverage: 10.01x in Q20 bases; sum-of-contigs Quality
              coverage: 9.91x in Q20 bases; agarose-1p
              -----
              * NOTE: This is a 'working draft' sequence. It currently
```

```
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 2365: contig of 2365 bp in length
* 2366 2465: gap of 100 bp
* 2466 37990: contig of 35525 bp in length
* 37991 90658: gap of 100 bp
* 90659 90759: gap of 100 bp
* 90759 102373: contig of 11615 bp in length
* 102374 102474: gap of 100 bp
* 102474 108662: contig of 6189 bp in length
* 108663 122646: contig of 13884 bp in length
* 122647 122747: gap of 100 bp
* 122747 143073: contig of 20327 bp in length
* 143074 143174: gap of 100 bp
* 143174 148249: contig of 5076 bp in length
* 148250 148350: gap of 100 bp
* 148350 162905: contig of 14556 bp in length.
FEATURES
             source
             1..162905
               /organism="Danio rerio"
               /mol_type="genomic DNA"
               /db_xref="taxon:7955"
               /clone="CH211-160J14"
               /clone_lib="CHORI-211"
             1..2365
               /notes="assembly fragment:00052
               fragment_chain:1"
             2466..37990
               /notes="assembly fragment:01265
               fragment_chain:1"
             38091..90658
               /notes="assembly fragment:01814
               fragment_chain:1"
             90759..102373
               /notes="assembly fragment:00273
               fragment_chain:1"
             102474..108662
               /notes="assembly fragment:00150
               fragment_chain:1"
             108763..122646
               /notes="assembly fragment:00441
               fragment_chain:1"
             122747..143073
               /notes="assembly fragment:00964
               fragment_chain:1"
             143174..148249
               /notes="assembly fragment:00088
               fragment_chain:1"
             148350..162905
               /notes="assembly fragment:00686
               fragment_chain:1
               clone_end:T7
               vector_side:right"
ORIGIN
Query Match      10.1%; Score 50.8; DB 2; Length 162905;
Best Local Similarity 51.3%; Pred. No. 0.05;
Matches 118; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 272  TAGTATTGATTTTACAGAGTTAATAAGTATATATATATATATATATATATATATATAT 331
DB 147965  TATATATATATATATATATTTTATATATATATATATATATATATATATATATATATA 148024

QY 332  TAATCTGTCGTTATCTAATTTAGTACGACCGACTCGTTGGCTAGGTCCTCT 391
DB 148025  AAAAAATATATATTTAATATATATATATATATATATATATATATATATAAAAAATTTAA 148084
```


----- Project Information
Center project name: zK154B17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 242361 bases at least Q40
Consensus quality: 242427 bases at least Q30
Consensus quality: 242459 bases at least Q20
Insert size: 242492; sum-of-contigs
Insert size: 240566; 2.5% error; agarose-fp
Quality coverage: 14.90x in Q20 bases; sum-of-contigs Quality
coverage: 15.02x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 242492: contig of 242492 bp in length.
Location/Qualifiers
1. .242492
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_lib="DKEY-154B17"
/clone_lib="DanioKey"
misc_feature 1. .242492
/note="assembly_fragment:03872
clone_end:SP6
vector_side:right"

ORIGIN
Query Match 10.1%; Score 50.6; DB 2; Length 242492;
Best Local Similarity 53.2%; Pred. No. 0.057;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 249 AGAGGAGAGATGGTTAGGCTTGTAGTTATTTTGTAGAGAGTTAATATAGTATATA 308
Db 165038 AAAATGTGATATTTTGTAGCAACATAAGAAATTTTGTATTTAAATGTATCATATAGATA 164979
QY 309 TATATATGATTAAGTGTACCTAATCTCTGCTGTTAATCTAATTTAGTGAGTCAGGAA 368
Db 164978 AATTTATTTCTATATTTATTTATTCGTAGACAGATTTGTACTTTAATTTAAGCAGTCAGATA 164919
QY 369 CCGACTCGTTGGCTAAGGCTCTCCTTTTGAAGCGCAACGTTCTTACTTTCATGTATATAA 428
Db 164918 TATAAAGTGTACCTTTAATTTTCTCTCTCAAGAAACACTCTTTTGAATGTATCAAA 164859
QY 429 ATACAGTCTGATCACAACA 449
Db 164858 CTAAGTACCTGCACAGACA 164838

RESULT 14
CR628325 264994 bp DNA linear HTG 03-OCT-2004
LOCUS
DEFINITION
Danio rerio clone DKEY-109N11, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION
CR628325
VERSION
CR628325.5 GI:53755980
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 264994)
McLaren,S.
AUTHORS
Direct Submission
TITLE

JOURNAL

COMMENT

Submitted (02-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfiah-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 3, 2004 this sequence version replaced gi:51127543.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfiah-help@sanger.ac.uk

----- Project Information

Center project name: zK109N11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 259248 bases at least Q40

Consensus quality: 260186 bases at least Q30

Consensus quality: 260917 bases at least Q20

Insert size: 263794; sum-of-contigs

Insert size: 258197; 3.9% error; agarose-fp

Quality coverage: 7.05x in Q20 bases; sum-of-contigs Quality
coverage: 7.20x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 27257: contig of 27257 bp in length
* 27258 27357: gap of 100 bp
* 27358 30685: contig of 3328 bp in length
* 30686 30785: gap of 100 bp
* 30786 63383: contig of 32598 bp in length
* 63384 70985: contig of 100 bp
* 70986 71085: gap of 100 bp
* 71086 90938: contig of 19853 bp in length
* 90939 91038: gap of 100 bp
* 91039 156734: contig of 65696 bp in length
* 156735 156834: gap of 100 bp
* 156835 189785: contig of 32951 bp in length
* 189786 189885: gap of 100 bp
* 189886 195030: contig of 5145 bp in length
* 195031 195130: gap of 100 bp
* 195131 220593: contig of 25463 bp in length
* 220594 220693: gap of 100 bp
* 220694 241129: contig of 20436 bp in length
* 241130 241229: gap of 100 bp
* 241230 245415: contig of 4186 bp in length
* 245416 245515: gap of 100 bp
* 245516 253885: contig of 8370 bp in length
* 253886 253985: gap of 100 bp
* 253986 264994: contig of 11009 bp in length.
Location/Qualifiers
1. .264994
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_lib="DKEY-109N11"
/clone_lib="DanioKey"
misc_feature 1. .27257
/note="assembly_fragment:00922
fragment_chain:1"
misc_feature 27358. .30685
/note="assembly_fragment:00042
fragment_chain:1"
misc_feature 30786. .63383
/note="assembly_fragment:01957
fragment_chain:1"
misc_feature 63484. .70985
/note="assembly_fragment:00135

```

misc_feature      fragment_chain:1"
71086. .90938
/notes="assembly_fragment:00658
fragment_chain:1"
misc_feature      91039. .156734
/notes="assembly_fragment:02393
fragment_chain:1"
misc_feature      156835. .189785
/notes="assembly_fragment:01540
fragment_chain:1"
misc_feature      189886. .195030
/notes="assembly_fragment:00078
fragment_chain:2"
misc_feature      195131. .220593
/notes="assembly_fragment:01225
fragment_chain:2"
misc_feature      220694. .241129
/notes="assembly_fragment:00412
fragment_chain:3"
misc_feature      241230. .245415
/notes="assembly_fragment:00007
fragment_chain:3"
misc_feature      245516. .253885
/notes="assembly_fragment:00210"
253986. .264994
/notes="assembly_fragment:00287.0"

ORIGIN
Query Match      10.1%; Score 50.6; DB 2; Length 264994;
Best Local Similarity 53.2%; Pred. No. 0.057; 94; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 249 AAGAGGAGAGATGGTTAGGGCTTAGTTATTGATTTTGTAGAGAGTTAATAATAGTATATA 308
Db 62333 AAAATGATATTTTAGACACATAAGAAATTTTGTATTAAGATGATCAGTTATAGATA 62392

Qy 309 TATATATGATATAGTATGTTACCTTAATCTGTCGTTAATCTAATTTAGTGTAGTCAGGAA 368
Db 62393 AATTATATTTCTATATTTATTCGTAAGACAGATTTGACTTTAATTTAAGCAGTCAGATA 62452

Qy 369 CCGACTGTTGGCTAAGTCTCTCTCTTTTGAACGCAAGCTTCTACTTCTATGATATATA 428
Db 62453 TATAAATGTTACCTTTAATTTTCTGCTCAAGAAACACTCTTTTGAATGATCAAA 62512

Qy 429 ATACAGTCTGATCACAAACA 449
Db 62513 CTAAAGTACCTGCACAGACA 62533

RESULT 15
CR450772
LOCUS      CR450772      277724 bp      DNA      linear      HTG 10-OCT-2004
DEFINITION      Danio rerio clone DKEY-1P24, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
ACCESSION      CR450772
VERSION      CR450772.2      GI:54021700
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE      1      (bases 1 to 277724)
AUTHORS      McLaren,S.
TITLE      Direct Submission
JOURNAL      Submitted (08-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 10, 2004 this sequence version replaced gi:47604430.
COMMENT      ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

```

```

Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zklp24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 275416 bases at least Q40
Consensus quality: 275745 bases at least Q30
Consensus quality: 275963 bases at least Q20
Insert size: 277024; sum-of-contigs
Insert size: 278944; 2.5% error; agarose-fp
Quality coverage: 5.97x in Q20 bases; sum-of-contigs Quality
coverage: 5.97x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 19819: contig of 19819 bp in length
* 19820 19919: gap of 100 bp
* 25522 25522: contig of 5603 bp in length
* 25523 25622: gap of 100 bp
* 25623 30418: contig of 4796 bp in length
* 30419 30518: gap of 100 bp
* 30519 196623: contig of 166105 bp in length
* 196624 196723: gap of 100 bp
* 196724 228104: contig of 31381 bp in length
* 228105 228204: gap of 100 bp
* 228205 262973: contig of 34769 bp in length
* 262974 263073: gap of 100 bp
* 263074 269787: contig of 6714 bp in length
* 269788 269887: gap of 100 bp
* 269888 277724: contig of 7837 bp in length.
*
FEATURES             Location/Qualifiers
     source            1. .277724
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone="DKEY-1P24"
                        /clone_lib="DanioKey"
     misc_feature      1. .19819
                        /notes="assembly_fragment:00205
                        fragment_chain:1"
     misc_feature      19920. .25522
                        /notes="assembly_fragment:00090
                        fragment_chain:1"
     misc_feature      25623. .30418
                        /notes="assembly_fragment:00009
                        fragment_chain:1"
     misc_feature      30519. .196623
                        /notes="assembly_fragment:00972
                        fragment_chain:1"
     misc_feature      196724. .228104
                        /notes="assembly_fragment:00401
                        fragment_chain:1"
     misc_feature      228205. .262973
                        /notes="assembly_fragment:00655
                        fragment_chain:1"
     misc_feature      263074. .269787
                        /notes="assembly_fragment:00043
                        fragment_chain:1"
     misc_feature      269888. .277724
                        /notes="assembly_fragment:00139.0"

ORIGIN
Query Match      10.1%; Score 50.6; DB 2; Length 277724;
Best Local Similarity 53.2%; Pred. No. 0.057; 94; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```

Qy	249	AAGAGGAGAGATGTTAGGCTTAGTTATTGATTTT	TAGAGAGTTAAATAAGTATATA	308
Db	85766	AAATGTGATATTTT	AGCAACATAGAAATTTTGTATTAATGTCAGTTATAGATA	85825
Qy	309	TATATATGTATAAGTAGGTTACCTAAATCTCTGTCGTTAAATCTAAATTTAGTGAGTCAGGAA	368	
Db	85826	AAATTTATTTCTATATATTATTCGTAAGACAGTATTGTACTTTAAATTTAAGCAGTCAGATA	85885	
Qy	369	CCGACTCGTTGGCTAAGGTCTCTCCTTTTGAACGGCAACGTTCTACTTTTCATGTATATAA	428	
Db	85886	TATAAACTGTACCTTTAAATTTTTCCTCTCAAGAAACACTCTTTTGAATGTATCAA	85945	
Qy	429	ATACAGTCTGATCACACA	449	
Db	85946	CTAAGTACCTGCACAGACA	85966	

Search completed: July 12, 2005, 05:59:52
Job time : 2540.52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 00:58:38 ; Search time 382.755 Seconds
(without alignments)
7748.542 Million cell updates/sec

Title: US-10-018-929C-2_COPY_6071_6571

Perfect score: 501

Sequence: 1 caactcatcagttactctc.....cttaaaaaaaaaaaaaa 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	6571	4	Aa89354 Arabidops
2	489.4	97.7	10329	4	Aa89385 Arabidops
3	489.4	97.7	10329	4	Aa89353 Arabidops
C 4	59.8	11.9	574	4	Aa89382 Brassica
5	44	8.8	102457	9	Continuation (4 of
6	43.4	8.7	6963	6	Abl32523 Human imm
C 7	42.6	8.5	416	8	Abx39419 Bovine ES
8	42.2	8.4	50899	12	Adq17045 Drosophil
C 9	41.8	8.3	7522	6	Abl32914 Human imm
C 10	41.4	8.3	475	10	Adb56428 Toxicity-
C 11	41.4	8.3	475	10	Adb50926 Primary r
12	41	8.2	1431	4	Abl01953 Drosophil
C 13	41	8.2	4414	4	Abl06106 Drosophil
14	41	8.2	5746	4	Abl01952 Drosophil
C 15	40.4	8.1	7167	6	Abl32400 Human imm
16	40.4	8.1	31241	10	Acc63515 Mycoplasma
17	40.4	8.1	31241	10	Acc69145 M. Genita
18	40.4	8.1	31241	12	Adn48950 Mycoplasma
19	40.4	8.1	110000	2	At58840 Mycoplasma
20	40.2	8.0	545	13	Acn53902 Cotton an

C 21	40.2	8.0	2314	8	Abx94632 C. elegan
C 22	40.2	8.0	2496	8	Abx94644 C. elegan
C 23	40.2	8.0	2497	8	Abx94641 C. elegan
C 24	40.2	8.0	2497	8	Abx94630 C. elegan
C 25	40.2	8.0	2497	8	Abx94640 C. elegan
C 26	40.2	8.0	2497	8	Abx94642 C. elegan
C 27	40.2	8.0	2497	8	Abx94646 C. elegan
C 28	40.2	8.0	2648	13	Adg91270 alpha-1 t
C 29	40.2	8.0	9260	8	Abx94633 C. elegan
C 30	40	8.0	17234	6	Abq67018 Human ang
C 31	39.6	7.9	5312	6	Abi32197 Human imm
C 32	39.6	7.9	7037	6	Abi33750 Human imm
C 33	39.6	7.9	7696	4	Aas46400 Tumour su
C 34	39.6	7.9	8900	13	Adg89411 Oligonuc
C 35	39.4	7.9	327	5	Abv08167 Human pro
C 36	39.2	7.8	1252	6	Abq54655 Human ova
C 37	39.2	7.8	8801	4	Aas45436 Chemical
C 38	39.2	7.8	8801	6	Abi33740 Human imm
C 39	39.2	7.8	8801	6	Abk28285 DNA trans
C 40	39.2	7.8	110000	4	Continuation (7 of
C 41	39.2	7.8	110000	4	Continuation (7 of
C 42	39.2	7.8	110000	6	Continuation (7 of
C 43	39.2	7.8	110000	6	Continuation (7 of
C 44	39.2	7.8	110000	12	ADH77486_06
C 45	39.2	7.8	170279	13	ABD32686_06

ALIGNMENTS

RESULT 1
AAA89354
ID AAA89354 standard; cDNA; 6571 BP.
XX
AC AAA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 310..6315
FT /*tag= a
XX
FN WO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP005761.
XX
PR 23-JUN-1999; 99GB-00014623.
XX
PA (NOVS) NOVARTIS AG.
XX
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GBS MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
DR P-PSDB; AAB20062.
XX
PT Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
XX
PS Claim 6; Page 22-32; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino acid
CC protein (see AAB20062) with similarity with ATPase/helicase proteins of
CC the SWI2/SNF2 family, which controls gene silencing, particularly plant

CC gene silencing. Trans-acting modified loci were identified by T-DNA
CC insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
CC carries a heritable inactivated, methylated hygromycin resistance gene
CC (hyg). Mutation of a silencing modified locus results in release of
CC silencing of the hyg gene and restores hygromycin resistance. Plants
CC homozygous for the silent resistance gene were subjected to
CC transformation with a bar selectable marker gene under control of the T-
CC DNA 1'-2', dual promoter. Transformants were selected and their progeny
CC screened for hygromycin resistance. The mutant phenotype (hygromycin
CC resistance) was screened for genetic co-segregation with a specific T-DNA
CC insert. Cloning of the tagged gene allowed characterization of the
CC present sequence. 2 Allelic differences were found in the corresponding
CC gene of A. thaliana ecotype Columbia (see AAA89385). Gene silencing is
CC useful as a molecular tool for regulating gene expression
XX
SQ Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 U; 0 Other;

Query Match 100.0%; Score 501; DB 4; Length 6571;
Best Local Similarity 100.0%; Pred. No. 1.6e-115;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACTCCATCAGTTACTCTCGCAAAATCAGGTCTCGGTTCTCTCGCACCACTAA 60
DB 6071 CAACTCCATCAGTTACTCTCGCAAAATCAGGTCTCGGTTCTCTCGCACCACTAA 6130
QY 61 ACTCATATAGACCATCTCTTCAACTCCGTCGCAGCTACTCCAACTCGTCAGTGC 120
DB 6131 ACTCATATAGACCATCTCTTCAACTCCGTCGCAGCTACTCCAACTCGTCAGTGC 6190
QY 121 CTCCTCAAGCTTTGACATATTCAGTCTGTCAATTCAGCAGCAGCAAGAACCAACCGC 180
DB 6191 CTCCTCAAGCTTTGACATATTCAGTCTGTCAATTCAGCAGCAGCAAGAACCAACCGC 6250
QY 181 AACAGACCTTGACAGTGGATTCAGAGCAACATGAAGTGGTTGTCTTCTGACGACG 240
DB 6251 AACAGACCTTGACAGTGGATTCAGAGCAACATGAAGTGGTTGTCTTCTGACGACG 6310
QY 241 AGTGACCTTAAGAGAGAGATGGTTAGGTCCTTACTTATTTAGAGAGTTAATAAT 300
DB 6311 AGTGACCTTAAGAGAGAGATGGTTAGGTCCTTACTTATTTAGAGAGTTAATAAT 6370
QY 301 AGTATATATATATATATAGTAGTGTACCTAATCTCTGTCGTTAATCTAATTTAGTGA 360
DB 6371 AGTATATATATATATATAGTAGTGTACCTAATCTCTGTCGTTAATCTAATTTAGTGA 6430
QY 361 GTACAGAACCGACTCGTTGGCTAAGTCTCTCTTTTGAACGCAACGTTCTACTTTTCAT 420
DB 6431 GTACAGAACCGACTCGTTGGCTAAGTCTCTCTTTTGAACGCAACGTTCTACTTTTCAT 6490
QY 421 GTATATAAATACAGTCTGATCACACACAAATTTGATTTGAAATATCTACTGATTTA 480
DB 6491 GTATATAAATACAGTCTGATCACACACAAATTTGATTTGAAATATCTACTGATTTA 6550
QY 481 ACTTAAAAAATAAAAAA 501
DB 6551 ACTTAAAAAATAAAAAA 6571

RESULT 2
AAA89385
ID AAA89385 standard; DNA; 10329 BP.
XX
AC AAA89385;
XX
DT 04-NOV-2004 (revised)
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana gene involved in epigenetic gene silencing.
XX
KW Gene silencing; silencing gene; MOM; ds.
XX
OS Arabidopsis thaliana.
XX

Key Location/Qualifiers
FH exon 1. .1008
FT /tag= a
FT /number= 1
FT intron 1009. .1295
FT /tag= b
FT /number= 1
FT exon 1296. .2550
FT /tag= c
FT /number= 2
FT CDS 1301. .9808
FT /tag= d
FT /product= "Columbia protein"
FT /note= "This sequence contains 15 introns."
FT intron 2551. .2673
FT /tag= e
FT /number= 2
FT exon 2674. .2752
FT /tag= f
FT /number= 3
FT intron 2753. .2867
FT /tag= g
FT /number= 3
FT exon 2868. .3113
FT /tag= h
FT intron 3114. .3506
FT /tag= i
FT /number= 4
FT exon 3507. .3680
FT /tag= j
FT /number= 5
FT intron 3681. .3973
FT /tag= k
FT /number= 5
FT exon 3974. .4895
FT /tag= l
FT /number= 6
FT allele 4338
FT /tag= m
FT /note= "ecotype Zurich"
FT intron 4896. .4975
FT /tag= n
FT /number= 6
FT exon 4976. .5217
FT /tag= o
FT /number= 7
FT intron 5218. .5777
FT /tag= p
FT /number= 7
FT exon 5778. .5882
FT /tag= q
FT /number= 8
FT intron 5883. .6082
FT /tag= r
FT /number= 8
FT exon 6083. .7480
FT /tag= s
FT /number= 9
FT allele 6721
FT /tag= t
FT /note= "ecotype Zurich"
FT intron 7481. .7615
FT /tag= u
FT /number= 9
FT exon 7616. .7771
FT /tag= v
FT /number= 10
FT intron 7772. .7914
FT /tag= w
FT /number= 10
FT exon 7915. .8070
FT /tag= x


```
FT /tag= d
FT /product= "MOM protein"
FT /note= "this sequence contains 15 introns."
FT 2551..2673
FT /tag= e
FT 2674..2752
FT /tag= f
FT /number= 3
FT 2753..2867
FT /tag= g
FT /number= 3
FT 2868..3113
FT /tag= h
FT /number= 4
FT 3114..3506
FT /tag= i
FT /number= 4
FT 3507..3680
FT /tag= j
FT /number= 5
FT 3681..3973
FT /tag= k
FT /number= 5
FT 3974..4895
FT /tag= l
FT /number= 6
FT 4338
FT /tag= m
FT /note= "ecotype Columbia"
FT 4896..4975
FT /tag= n
FT /number= 6
FT 4976..5217
FT /tag= o
FT /number= 7
FT 5218..5777
FT /tag= p
FT /number= 7
FT 5778..5882
FT /tag= q
FT /number= 8
FT 5883..6082
FT /tag= r
FT /number= 8
FT 6083..7480
FT /tag= s
FT /number= 9
FT 6721
FT /tag= t
FT /note= "ecotype Columbia"
FT 7481..7615
FT /tag= u
FT /number= 9
FT 7616..7771
FT /tag= v
FT /number= 10
FT 7772..7914
FT /tag= w
FT /number= 10
FT 7915..8070
FT /tag= x
FT /number= 11
FT 8071..8153
FT /tag= y
FT /number= 11
FT 8154..8318
FT /tag= z
FT /number= 12
FT 8319..8451
FT /tag= aa
FT /number= 12
FT 8452..8629
```

```
FT intron /tag= ab
FT /number= 13
FT 8630..8718
FT /tag= ac
FT /number= 13
FT 8719..8918
FT /tag= ad
FT /number= 14
FT 8919..9000
FT /tag= ae
FT /number= 14
FT 9001..9211
FT /tag= af
FT /number= 15
FT 9212..9284
FT /tag= ag
FT /number= 15
FT 9285..10329
FT /tag= ah
FT /number= 16
FT WO200100801-A2.
FT 04-JAN-2001.
FT 21-JUN-2000; 2000WO-EP005761.
FT 23-JUN-1999; 99GB-00014623.
FT (NOVS ) NOVARTIS AG.
FT (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
FT WPI; 2001-137952/14.
FT P-PSDB; AAB20062.
FT Novel gene encoding a protein that controls gene silencing, in particular
FT silencing of plant genes.
FT Claim 6; Page 19-22; 48pp; English.
FT
FT The present sequence is that of an Arabidopsis thaliana gene (MOM)
FT involved in epigenetic gene silencing. It encodes a 2001-amino acid
FT protein (see AAB20062) with similarity with ATPase/helicase proteins of
FT the SWI2/SNF2 family, which controls gene silencing, particularly plant
FT gene silencing. Trans-acting modified loci were identified by T-DNA
FT insertion mutagenesis using Arabidopsis thaliana ecotype Zurich, which
FT carries a heritable inactivated, methylated hygromycin resistance gene
FT (hyg). Mutation of a silencing modified locus results in release of
FT silencing of the hyg gene and restores hygromycin resistance. Plants
FT homozygous for the silent resistance gene were subjected to
FT transformation with a bar selectable marker gene under control of the T-
FT DNA 1'-2' dual promoter. Transforms were selected and their progeny
FT screened for hygromycin resistance. The mutant phenotype (hygromycin
FT resistance) was screened for genetic co-segregation with a specific T-DNA
FT insert. Cloning of the tagged gene allowed characterization of the
FT present sequence. 2 Allelic differences were found in the corresponding
FT gene of A. thaliana ecotype Columbia, occurring at nucleotide positions
FT 4338 (A instead of T) and 6721 (T instead of G). Gene silencing is useful
FT as a molecular tool for regulating gene expression
FT
FT Revised record issued on 04-NOV-2004 : Correction to feature table key
FT
FT Sequence 10329 BP; 3105 A; 1934 C; 2142 G; 3148 T; 0 U; 0 Other;
FT
FT Query Match 97.7%; Score 489.4; DB 4; Length 10329;
FT Best Local Similarity 98.8%; Pred. No. 1.5e-112;
FT Matches 493; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
FT
FT Qy 1 CAACTCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGCTTCTCTGCACCACTAA 60
FT |||||
FT Db 9564 CAACTCCATCAGTTACTCTCTGCAACAAATCCAGGTCTGCTTCTCTGCACCACTAA 9623
```



```

XX PN WO200200928-A2.
XX PA
XX PD 03-JAN-2002.
XX PF
XX PJ 02-JUL-2001; 2001WO-EP007537.
XX PP
XX PR 30-JUN-2000; 2000DE-01032529.
XX PS 01-SEP-2000; 2000DE-01043826.
XX PT
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX PJ WPI; 2002-130909/17.
XX PP
XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.
XX PS Claim 1; SEQ ID NO 496; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention
XX SQ Sequence 6963 BP; 1650 A; 142 C; 1711 G; 3460 T; 0 U; 0 Other;

Query Match 8.7%; Score 43.4; DB 6; Length 6963;
Best Local Similarity 53.9%; Pred. No. 1.3;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 186 AGCTTGACGAGTGATGTCGAGACCAACATGAAGTGTGTTCTCTGACGAGCTGA 245
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4047 AGTTTGGTAGGTAATTAGAGAGATGAGATTCGGGTTGTAATTTGTTTGTAGGGT 4106
QY 246 CCTAAGAGGAGAGATGTTAGGGTCTTAGTTATTGATTTTATAGAGCTTAATAATAGTAT 305
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4107 TTTACGAGAGATAAGGTAGAGAAATTTTITTTTTTTTTTTTATATATATATAT 4166
QY 306 ATATATATATGTAATAGTAGGTACCTAACTCTGTCGTTAACT 350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4167 ATATATATATATATATATATATATTTTCGTTAATAGGTTAGATTATAT 4211

RESULT 7
ABX39419/c
ID ABX39419 standard; cDNA; 416 BP.
XX AC
XX AC ABX39419;
XX AC
XX DT 20-FEB-2003 (first entry)
XX DE
XX DE Bovine EST associated with lactation/muscle/fat deposition #4584.
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
XX KW muscle deposition; fat deposition; genome mapping; gene identification;
XX KW gene analysis; cattle breeding.
XX OS Bos Taurus.
XX PN US2002137139-A1.
XX PD
XX PD 26-SEP-2002.
XX PF
XX PF 24-SEP-2001; 2001US-00960352.
XX PP
XX PP 12-JAN-1999; 99US-0115707P.

```

```

PR 11-JAN-2000; 2000US-00480902.
XX (BYAT/) BYATT J C.
XX PA (MATH/) MATHIALAGAN N.
XX PA (TAON/) TAO N.
XX PA (WARR/) WARREN W C.
XX
XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX PJ WPI; 2003-110599/10.
XX PP
XX PT New nucleic acid associated with lactation, and muscle and fat
XX PT deposition, useful for genome mapping, gene identification and analysis,
XX PT cattle breeding, or for genetically improving cattle.
XX PS Claim 2; SEQ ID NO 4584; 245pp; English.
XX CC The invention relates to a purified nucleic acid molecule associated with
XX CC lactation or muscle and fat deposition (designated LMPD), derived from
XX CC cattle, and the LMPD nucleic acid can specifically hybridize to a second
XX CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
XX CC appearing as ABX34836-ABX49947, or complements of them. Also included are
XX CC ; (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
XX CC acid linked to a promoter and a 3' non-translated sequence that
XX CC functions in the cell to cause termination of transcription and addition
XX CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
XX CC (2) determining a level or pattern of a molecule in a bovine cell or
XX CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
XX CC of the 15112 nucleic acid sequences or its complement or fragment) with a
XX CC complementary nucleic acid molecule obtained from the bovine cell or
XX CC tissue, where hybridisation between the marker nucleic acid and the
XX CC complementary nucleic acid permits the detection of the molecule; and (b)
XX CC detecting the level or pattern of the complementary nucleic acid, where
XX CC the detection of the complementary nucleic acid is predictive of the
XX CC level or pattern of the molecule. The LMPD nucleic acid is used for
XX CC determining a level or pattern of a molecule in a bovine cell or tissue.
XX CC It is useful for genome mapping, gene identification and analysis, cattle
XX CC breeding, preparation of constructs for use in cattle gene expression, or
XX CC for genetically improving cattle. The present sequence is one of the
XX CC 15112 bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The
XX CC present sequence was not shown in the specification but was obtained in
XX CC electronic format from the USPTO web site:
XX CC seqdata.uspto.gov/sequence.html?DocID=20020137139
XX SQ Sequence 416 BP; 223 A; 1 C; 14 G; 178 T; 0 U; 0 Other;

Query Match 8.5%; Score 42.6; DB 8; Length 416;
Best Local Similarity 48.9%; Pred. No. 0.9;
Matches 114; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 269 TCTTAGTTATTGATTTTATAGAGAGTTAATAATAGTATATATATATATATAGTAGGTT 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 TTTTITTTTTTATGAAATTTTAAATTTTATATATTTTAAATTTTAAATCTTTTTTTTTT 174
QY 329 ACCTAATCTCTGTCGTTAATCTAATTTAGTGAGTCAGGACCGACGTCGTTGGCTAAGGTC 388
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 TTTTITTTTTTTTTTTTTTAAATTTTATTTTATTTTATATAAAATTTATTTTATTATT 114
QY 389 TCTCCTTTGAAACGCAACGTTCTTCTTTCATGATATATAAATACAGTCGTCATCACAAAC 448
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 TTTCTATTTTTCATTTTAAATTTTTTTTTTTTTTAAAAAATAAAAAAATACCAAAAC 54
QY 449 ACAAAATTGATGTTGAAATACTACTGATTTAACTTAAAAAATAAAAAAATAAAAAA 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 CAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1

RESULT 8
ADQ17045
ID ADQ17045 standard; DNA; 50899 BP.
XX AC
XX AC ADQ17045;
XX

```



```
Db      844 AGATTTTAAATGTTTGGTACTTAACRAATTTAAATATATTAGTATTAGATATA 785
Qy      317 TATAAGTAGGTACCTAATCTCTGTCGTTTAACTAATTTAGTAGTCAGGACGACTCG 376
Db      784 CATTGTGTAGTTTATTTAAACTTTTATTTTTCATTTTTCATTTAGTTTGTCTTGAATCGCTGG 725
Qy      377 T 377
Db      724 T 724

RESULT 14
ABL01952
ID   ABL01952 standard; cDNA; 5746 BP.
XX
AC   ABL01952;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 338.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
DR   P-PSDB; ABB57849.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 338; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5746 BP; 1776 A; 1219 C; 1123 G; 1628 T; 0 U; 0 Other;

Query Match      8.2%; Score 41; DB 4; Length 5746;
Best Local Similarity 50.2%; Pred. No. 5.1;
Matches 129; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

Qy      247 CTAAGAGGAGATGGTTAGGCTTCTAGTTATTGATTTTAGAGAGTTAATAATAGTATA 306
Db      4469 CCATGAGAAAGCTTGTTATGTTTATTTCTGTTTCTAAAGTATATAAAATGATTTA 4528
Qy      307 TATATATATGATTAAGTAGGTTACCTAATCTCT---GTCGTTAATCTAATTTAGTAGTC 363
Db      4529 AATATATAAATATATATATAATAAACTTAGTTCTTATAGTCGTGACCTAAATATAACTCGCC 4588
```

```
Qy      364 AGGAACCGACTCGTTGGCTAAGGTCTCTCTCTTTTGAACCGCAACGTTCTACTTTCTATGTA 423
Db      4589 AGCAATTTGTATATATAGTATGATAGCGGAATTTGATAGCAACGAGAAACACAGCTA 4648
Qy      424 TATAAATACAGTCTGATCACAACACAAATTCGATGATTGAAAATACTACTGATTTAACT 483
Db      4649 AAAACTAACATTTAGTTTACAAAAAATTATAAAATACAATTTAAAGATGACGAAATCAAT 4708
Qy      484 TAAAAAATAAAAAA 500
Db      4709 TAAAAAAGAAAGCAAA 4725

RESULT 15
ABL32400/c
ID   ABL32400 standard; DNA; 7167 BP.
XX
AC   ABL32400;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 373.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 373; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 7167 BP; 2411 A; 47 C; 1427 G; 3282 T; 0 U; 0 Other;

Query Match      8.1%; Score 40.4; DB 6; Length 7167;
Best Local Similarity 48.7%; Pred. No. 7.7;
Matches 110; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy      276 TATTGATTTTAGAGAGTTAATAATAGTATATATATATATATATAGTTACGTTACCTAAT 335
```


THIS PAGE BLANK (USPTO)


```
/ LENGTH: 3344
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
/ OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3504

Query Match          7.5%; Score 37.8; DB 4; Length 3344;
Best Local Similarity 54.7%; Pred. No. 1.2; Mismatches 62; Indels 0; Gaps 0;
Matches 75; Conservative 0;

QY 342 CGTTAATCTAATTAGTAGAGTCAGAACCGACTCGTTGGCTAAGGTCTCTCCTTTTGA 401
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 296 CTTTATTGCTTTGATTGCACTAACATCTAGATTGTTGCTAAACGTTGTTGAAGAA 237
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 CGCAACGCTTCTACTTTTCATGTATATAAATACAGTCTGATCACACACAAATTTGAT 461
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 236 CATACCAATCTATCATCAAAATATTTAAATAAATTTATGATATAATGATAAAAGAGGTGA 177
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 TGAATAACTACTGATT 478
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 176 TAAAAATGAAAATGATT 160
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-09-949-016-14120/c
; Sequence 14120, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 14120
; LENGTH: 125192
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(125192)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14120

Query Match          7.5%; Score 37.6; DB 4; Length 125192;
Best Local Similarity 49.0%; Pred. No. 8.9;
Matches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 295 AATAATAGTATATATATATATATATATATATATATATATATATATATATATATAT 354
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64575 AGTGACTGTAGTTAAATTTATATATATATATATATATATATATATATATATAT 64516
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 355 TAGTGAGTCAGAACCGACTCGTTGGCTCTCTCCTTTTGAACGCAACGCTCTAC 414
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64515 TAAATGTTCTTACCACAAAGCAATGATAAATTTTGAAGTGATGGAATGCTACTATTC 64456
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 415 TTTTCATGTATATAATACAGTCTGATCACACACAAATTTGATGTTGAAATTAATCTACT 474
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64455 TGACTTGTATACATATATCAAAAGATCACAGTATACCCCAATATAAATTAATTA 64396
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 GATTTAACCTTAAAAAATAAAAA 498
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64395 TTGTCATTAATAAATAAAAAA 64372
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13

```
US-09-949-016-17362
; Sequence 17362, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 17362
; LENGTH: 134434
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION:
US-09-949-016-17362
```

```
Query Match          7.5%; Score 37.4; DB 4; Length 134434;
Best Local Similarity 58.8%; Pred. No. 11;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 276 TATTGATTTTACAGAGTTAAATAAGTATATATATATATATATATATATATATATAT 335
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 34681 TATACATTTTAAATATTTTATATATATATATATATATATATATATATATATAT 34740
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 336 CTCTGTCGTTAACTAATTTAGTGTACAGAACCGACTCGTTGGCTAAGG 386
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 34741 TTTTNTTTTTTTTTTTTTTTTTTTTGTGACACAGAGCTTGCTCTGTCAACCCAGG 34791
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 14

```
US-09-949-016-16600/c
; Sequence 16600, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 16600
; LENGTH: 298336
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(298336)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16600
```

```
Query Match          7.5%; Score 37.4; DB 4; Length 298336;
Best Local Similarity 47.0%; Pred. No. 16;
Matches 116; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```

QY 255 AGAGATGGTTCAGTCTAGTATTGATTTTTCAGAGAGTTAAATAGTATATATATA 314
Db 274224 ATATAATTTATATATTTTATATATATATATATATATATATATATATATA 274165
QY 315 TGTATAAGTAGTGGTACCTAATCTCTGTCGTTAACTAATTTAGTGAGTCAGGAACCGACT 374
Db 274164 AATTTATATATTTTA 274105
QY 375 CGTTGGCTAAGGTCCTCTCTTTTGAAGCGCAACGTTCTACTTTTCATGTATATAAATACAG 434
Db 274104 TATTATACATGTA 274045
QY 435 TCATGATCACACACACAAATGATGATGAAATACACTGACTGATTTAACTTAAAAA 494
Db 274044 TAAATATATATTTA 273985
QY 495 AAAAAAA 501
Db 273984 TATAATA 273978

RESULT 15
US-09-949-016-81754/c
; Sequence 81754. Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 81754
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-81754

Query Match 7.4%; Score 37.2; DB 4; Length 601;
Best Local Similarity 48.5%; Pred. No. 0.77; 104; Indels 0; Gaps 0;
Matches 99; Conservative 1; Mismatches 104; Indels 0; Gaps 0;
QY 295 AATAATAGTATATATATATATATAGTAGGTTACCTAATCTCTGTCGTTAAATCTAATT 354
Db 444 AGTGACTGTAGTTAAATTTATATATATATATATATATATATATATATATATATATAT 385
QY 355 TAGTGAGTCAGGAACCGACTCGTTGGCTAGGTCCTCTCTTTTGAAGCGCAACGTTCTAC 414
Db 384 TAAATGTTCTTACCACAAAGCAATGATAAATATTTGAGGTGATGGATATGCTACTATTC 325
QY 415 TTTCATGTATATAATACAGTCTGATCACACACAAATGATGATTTGAAATACTACT 474
Db 324 TGACTTGTATACATATATCAAAAGATCAGATATACCCCATATAAATATAAACAATTATA 265
QY 475 GATTTAACTTAAAAA 498
Db 264 TTGTCATTAATAAATAAATAA 241

Search completed: July 12, 2005, 10:18:13
Job time : 118.576 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:07:39 ; Search time 543.528 Seconds
(without alignments)
5787.078 Million cell updates/sec

Title: US-10-018-929c-2_COPY_6071_6571

Perfect score: 501

Sequence: 1 caactcatcagttactctc.....cttaaaaaaaaaaaaaaaaaa 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6330945 seqs, 3139162390 residues

Total number of hits satisfying chosen parameters: 12661890

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	8.8	402850	10	US-09-844-653-5 Sequence 5, Appli
2	43.4	8.7	6963	15	US-10-311-455-496 Sequence 496, App
3	42.6	8.5	416	9	US-09-960-352-4584 Sequence 4584, Ap
4	42.2	8.4	50899	19	US-10-630-660-10 Sequence 10, Appl
5	42	8.4	286	18	US-10-424-599-85690 Sequence 85690, A
6	41.8	8.3	7522	15	US-10-311-455-887 Sequence 887, App
7	41.8	8.3	3673778	16	US-10-312-841-1 Sequence 1, Appli

c	8	41.6	8.3	1444	19	US-10-437-963-74011	Sequence 74011, A
c	9	40.4	8.1	7167	15	US-10-311-455-373	Sequence 373, App
	10	40.4	8.1	31241	10	US-09-960-870-13	Sequence 13, Appl
	11	40.4	8.1	31241	10	US-09-960-858-13	Sequence 13, Appl
	12	40.4	8.1	31241	18	US-10-251-668-13	Sequence 13, Appl
	13	40.4	8.1	580073	16	US-10-205-220-1	Sequence 1, Appli
	14	40.2	8.0	545	19	US-10-021-323-8683	Sequence 8683, Ap
	15	40.2	8.0	2314	21	US-10-481-582-6	Sequence 6, Appli
c	16	40.2	8.0	6675	20	US-10-481-582-7	Sequence 7, Appli
c	17	40.2	8.0	30535	21	US-10-451-503A-7	Sequence 7, Appli
c	18	40	8.0	606	20	US-10-425-115-9936	Sequence 9936, Ap
c	19	40	8.0	17234	19	US-10-433-793-48	Sequence 48, Appl
c	20	39.6	7.9	294	18	US-10-424-599-72610	Sequence 72610, A
c	21	39.6	7.9	440	20	US-10-425-115-2123	Sequence 2123, Ap
c	22	39.6	7.9	5312	15	US-10-311-455-170	Sequence 170, App
c	23	39.6	7.9	7037	15	US-10-311-455-1723	Sequence 1723, Ap
c	24	39.6	7.9	7696	18	US-10-221-714A-122	Sequence 122, App
c	25	39.6	7.9	3673778	16	US-10-312-841-1	Sequence 1, Appli
c	26	39.4	7.9	327	20	US-10-357-930-8158	Sequence 8158, Ap
c	27	39.4	7.9	3908	20	US-10-425-115-70814	Sequence 70814, A
c	28	39.2	7.8	1252	17	US-10-264-049-535	Sequence 535, App
c	29	39.2	7.8	8801	14	US-10-239-676-143	Sequence 143, App
c	30	39.2	7.8	8801	15	US-10-311-455-1713	Sequence 1713, Ap
c	31	39.2	7.8	8801	15	US-10-240-453-159	Sequence 159, App
c	32	39.2	7.8	170279	19	US-10-388-838-1	Sequence 1, Appli
c	33	39.2	7.8	1503841	9	US-09-795-668-1	Sequence 1, Appli
c	34	39.2	7.8	1503841	9	US-09-795-686-1	Sequence 1, Appli
c	35	39.2	7.8	1503841	9	US-09-946-807-1	Sequence 1, Appli
c	36	39	7.8	7676	15	US-10-240-485-151	Sequence 151, App
c	37	39	7.8	17389	15	US-10-311-455-1388	Sequence 1388, Ap
c	38	39	7.8	56153	18	US-10-221-714A-519	Sequence 519, App
c	39	39	7.8	83391	19	US-10-433-793-124	Sequence 124, App
c	40	39	7.8	271990	15	US-10-195-144-87	Sequence 87, Appl
c	41	39	7.8	271990	17	US-10-345-072-87	Sequence 87, Appl
c	42	38.8	7.7	1054	15	US-10-180-375-125	Sequence 109, App
c	43	38.8	7.7	1054	17	US-10-183-687-125	Sequence 125, App
c	44	38.6	7.7	893	20	US-10-425-115-82551	Sequence 82551, A
c	45	38.6	7.7	1649	19	US-10-767-795-5688	Sequence 5688, Ap

ALIGNMENTS

RESULT 1
US-09-844-653-5
; Sequence 5, Application US/09844653
; Publication No. US20030054347A1
; GENERAL INFORMATION:
; APPLICANT: Richards, Julia
; APPLICANT: Rozsa, Frank
; TITLE OF INVENTION: Detecting and Treating Eye Disease
; FILE REFERENCE: UM-06105
; CURRENT APPLICATION NUMBER: US/09/844,653
; CURRENT FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 402850
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (46565)..(46565)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (117443)..(118256)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (118272)..(118736)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (120172)..(120640)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.

```

; NAME/KEY: misc feature
; LOCATION: (122654)..(122654)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (203527)..(203963)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (215340)..(215340)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (220202)..(220650)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc feature
; LOCATION: (222415)..(222416)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; US-09-844-653-5

Query Match      8.8%; Score 44; DB 10; Length 402850;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 110; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

Qy 275 TTATTGATTTTGTAGAGAGTTAAATAGTAGTATATATATATATATATAGTAGGTACCTAA 334
Db 311241 TTATATTATATAAATACATTATATATAAATATATATATATATATATATATATATAT 311300

Qy 335 TCCTCTGTCGTTAACTCTAAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAAGGTCCTCCT 394
Db 311301 TATACATATATATATAAATACATGAGTGTGCTGAGATGCACCTGTTTCTCAAGTCCTTAT 311360

Qy 395 TTTGAAACGCAACGCTTCTACTTTTCATGTATATATAATACAGTCGTGATCACACACAAAT 454
Db 311361 GTCTTTTGTTCAGTACTCTTTTATAAATAATATATGATATAGATGATATACATATGATTGAAT 311420

Qy 455 TGATGATTGAAATACTACTGATTTTAACTTAAAAA 494
Db 311421 TCATTTATAAAGAAATGCGCATATCTTAAAAAATTTAA 311460

RESULT 2
US-10-311-455-496
; Sequence 496, Application US/10311455
; Publication No. US2003014306A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Expression of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043026.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 496
; LENGTH: 6963
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-311-455-496

Query Match      8.7%; Score 43.4; DB 15; Length 6963;
Best Local Similarity 53.9%; Pred. No. 1.3;
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 186 AGCTTGACGAGTGATGTCGAGACCAACAAATGAAGTGTGCTTTCTGACGACGAGTGA 245
Db 4047 AGGTTTGGTTAGGTATTAGAGAGATAGGATTTCCGGGTTGTATTCTTTTAGTAGGGT 4106
```

```

Qy 246 CCTAAGAGAGAGATGCTTAGGCTCTTAGTATTGATTTTGTAGAGAGTTAATAATAGTAT 305
Db 4107 TTTACGAGAGGATAGGTTAGAGAAATTTTTTTTTTTTTTTTTTATATATATATATAT 4166

Qy 306 ATATATATATGTATAAGTAGGTGTACCTAAATCTCTGCTGCTGTTAAATCT 350
Db 4167 ATATATATATATATATATATATTTTCGTTAATAGGTTAGATTATAT 4211

RESULT 3
US-09-960-352-4584/c
; Sequence 4584, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND LACTATION-INDUCED MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4584
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB3057-016-Q1-K1-E11
; US-09-960-352-4584

Query Match      8.5%; Score 42.6; DB 9; Length 416;
Best Local Similarity 48.9%; Pred. No. 0.55;
Matches 114; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 269 TCTTAGTTATTGATTTTGTAGAGAGTTAAATAATAGTAGTATATATATATATATATAGTAGGT 328
Db 233 TTTTITTTTATGCAATTTTAAATTTTATATATTTTAAATCTTTTAAATCTTTTITTTT 174

Qy 329 ACCTAATCTCTGCTGTTAAATCTAATTTAGTAGTCAGGAACCGACTGCTGGCTAGGTC 388
Db 173 TTTTITTTTITTTTAAATTTTATTTTITTTTATATAAAATTTTATTTTATTTATTT 114

Qy 389 TCTCCTTTTGAACGCAACGCTCTACTTTTCATGTATATAATACAGTCCTCATCACCAAC 448
Db 113 TTCTATTTTCAATTTAAATTTTITTTTAAAAAATTTAAAAAATACCAAAAC 54

Qy 449 ACAATTTGATTTGAAATATACTACTGATTTAACTTAAAAAATTTAAAAAATTTAAAAA 501
Db 53 CAAAAAATAAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 1

RESULT 4
US-10-630-660-10
; Sequence 10, Application US/10630660
; Publication No. US2004012744A1
; GENERAL INFORMATION:
; APPLICANT: Spradling, Allan C.
; APPLICANT: Stewart, Dianne R.
; APPLICANT: Tulin, Alexei V.
; TITLE OF INVENTION: Poly-(ADP) ribose polymerase enzyme and uses thereof
; FILE REFERENCE: 56100-5031-US
; CURRENT APPLICATION NUMBER: US/10/630,660
; CURRENT FILING DATE: 2003-07-31
; PRIOR APPLICATION NUMBER: US 60/399,460
; PRIOR FILING DATE: 2002-07-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 50899
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
```


Db 5811 ATCGCCTTTGGTCTATTCTCATATTGATTAGTGTCTTAGGAGTCTGTTTAAATACCCCTG 5870
Qy 448 CACAAATGTAGTGAATAAATCTACTGATTTAACTTAACTTAAATAAATAAATAAATAAATAA 501
Db 5871 CAAAAGTATTTAAGGGGAAAAAGAAAGGAAATAATCAAAATAAGACAATGAAAA 5924

RESULT 11

US-09-960-858-13
; Sequence 13, Application US/09960858
; Publication No. US20030138777A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
; FILE REFERENCE: P-EA 4974
; CURRENT APPLICATION NUMBER: US/09/960,858
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31241
; TYPE: DNA
; ORGANISM: M. genitalium
US-09-960-858-13

Query Match 8.1%; Score 40.4; DB 10; Length 31241;
Best Local Similarity 48.3%; Pred. No. 16;
Matches 113; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 268 GTCTTAGTATTGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATATAGTAGGT 327
Db 5691 GTACTAGTAGTATTGCAACTAGTTTGTGATGATTTATGTAATTAACCTCTTTAGTACAAGGA 5750
Qy 328 TACCTAATCTCTGCTGTTAATCTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAAGGT 387
Db 5751 TCAGTAAAACTATAGCACTGAACTCTATTCAATTAGAAAAGGAATTAAGCATGGCA 5810
Qy 388 CTCTCCCTTTGAAACGCAACGTTCTACTTTTCATGTATATAATAACAGTCTGATCACAA 447
Db 5811 ATCGCCTTTGGTCTATTCTCATATTGATTAGTGTCTTAGGAGTCTGTTTAAATACCCCTG 5870
Qy 448 CACAAATGTAGTGAATAAATCTACTGATTTAACTTAACTTAAATAAATAAATAAATAAATAA 501
Db 5871 CAAAAGTATTTAAGGGGAAAAAGAAAGGAAATAATCAAAATAAGACAATGAAAA 5924

RESULT 12

US-10-251-668-13
; Sequence 13, Application US/10251668
; Publication No. US20040063097A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
; FILE REFERENCE: P-EA 5441
; CURRENT APPLICATION NUMBER: US/10/251,668
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 09/960,607
; PRIOR FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31241
; TYPE: DNA
; ORGANISM: M. genitalium
US-10-251-668-13

Query Match 8.1%; Score 40.4; DB 18; Length 31241;
Best Local Similarity 48.3%; Pred. No. 16;
Matches 113; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 268 GTCTTAGTATTGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATATAGTAAAGTAGGT 327

Db 5691 GTACTAGTAGTATTGCAACTAGTTTGTGATGATTATGTAATTAACCTCTTTAGTACAAGGA 5750
Qy 328 TACCTAATCTCTGCTGTTAATCTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAAGGT 387
Db 5751 TCAGTAAAACTATAGCAACTGAACTCTATTCAATTAGAAAAGGAATTAAGCATGGCA 5810
Qy 388 CTCTCCCTTTGAAACGCAACGTTCTACTTTTCATGTATATAATAACAGTCTGATCACAA 447
Db 5811 ATCGCCTTTGGTCTATTCTCATATTGATTAGTGTCTTAGGAGTCTGTTTAAATACCCCTG 5870
Qy 448 CACAAATGTAGTGAATAAATCTACTGATTTAACTTAAATAAATAAATAAATAAATAAATAA 501
Db 5871 CAAAAGTATTTAAGGGGAAAAAGAAAGGAATAATCAAAATAAGACAATGAAAA 5924

RESULT 13

US-10-205-220-1
; Sequence 1, Application US/10205220
; Publication No. US20030170663A1
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; FILE REFERENCE: PB193P1D1
; CURRENT APPLICATION NUMBER: US/10/205,220
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 08/545,528
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-10-205-220-1

Query Match 8.1%; Score 40.4; DB 16; Length 580073;
Best Local Similarity 48.3%; Pred. No. 66;
Matches 113; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy 268 GTCTTAGTATTGATTTTGTAGAGAGTTAATAATAGTATATATATATATATATATATAGTAGGT 327
Db 52978 GTACTAGTAGTATTGCAACTAGTTTGTGATGATTATGTAATTAACCTCTTTAGTACAAGGA 53037
Qy 328 TACCTAATCTCTGCTGTTAATCTAATTTAGTGAGTCAGGAACCGACTCGTTGGCTAAAGGT 387
Db 53038 TCAGTAAAACTATAGCAACTGAACTCTATTCAATTAGAAAAGGAATTAAGCATGGCA 53097
Qy 388 CTCTCCCTTTGAAACGCAACGTTCTACTTTTCATGTATATAATAACAGTCTGATCACAA 447
Db 53098 ATCGCCTTTGGTCTATTCTCATATTGATTAGTGTCTTAGGAGTCTGTTTAAATACCCCTG 53157
Qy 448 CACAAATGTAGTGAATAAATCTACTGATTTAACTTAAATAAATAAATAAATAAATAAATAA 501
Db 53158 CAAAAGTATTTAAGGGGAAAAAGAAAGGAATAATCAAAATAAGACAATGAAAA 53211

RESULT 14

US-10-021-323-8683
; Sequence 8683, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C. C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; Plants

FILE REFERENCE: 38-21(52274)B
CURRENT APPLICATION NUMBER: US/10/021.323
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: US 60/255, 619
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 17880
SEQ ID NO 8683
LENGTH: 545
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(545)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3828-014-Q1-K6-H10
US-10-021-323-8683

Query Match 8.0%; Score 40.2; DB 19; Length 545;
Best Local Similarity 47.8%; Pred. No. 2.7;
Matches 117; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY	257	AGATGGTTAGGGCTTGTAGTTATGATTTTGTAGAGAGTTAATAATAGTATATATATATATG	316
Db	254	AGATGCTGTGGTGAATTGGTTAGTATCTTTTATTTGCTTATATTTGGTATATTACCATATG	313
QY	317	TATAAGTAGGTTACCTAATCTCTCGTCTAATCTAATTTAGTGTAGTCAGGAACGACTCG	376
Db	314	AAAAAGAGTATTCGATCGTCTCGGGTTAGTTCGGATTTAAACAGCAATCCAGGGTTG	373
QY	377	TTGGCTAAGGCTCTCTCCTTTTGAACGCAACGGTTCTACTTTTCATGTATATAAATACAGTC	436
Db	374	GTATTATTAGTATTATGATGAAGCAACTACTTTTGAGAGTTCAAAAAAATAAAAAA	433
QY	437	TGATCACACACACAAATTCATGATTCGAAATACTACTGATTTAACTTAAAAAATAAAAA	496
Db	434	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	493
QY	497	AAAAA 501	
Db	494	AAAAA 498	

RESULT 15
US-10-481-582-6/c
Sequence 6, Application US/10481582
Publication No. US20050054592A1
GENERAL INFORMATION:
APPLICANT: BAUMEISTER, Ralf
TITLE OF INVENTION: IDENTIFICATION OF SES-1 AND THE USES OF SAME
FILE REFERENCE: DEAV2001/0082 US PCT
CURRENT APPLICATION NUMBER: US/10/481.582
CURRENT FILING DATE: 2003-12-22
PRIOR APPLICATION NUMBER: DE 10130166.9
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: PCT/EP02/06965
PRIOR FILING DATE: 2002-06-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 2314
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-10-481-582-6

Query Match 8.0%; Score 40.2; DB 21; Length 2314;
Best Local Similarity 53.5%; Pred. No. 5.4;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY	345	TAATCTAATTTAGTGAACCGACTCGTTGGCTAAGGCTCTCTCCTTTTGAACGC	404
Db	2099	TAATATTATTAGAAATTTACTAACTATGAGCGAGCTATTTTGACGGGGATGGGAAGG	2040

QY	405	AACGTTCTACTTTTCATGTATATAAATACAGTCTGTATCACACACAAATTTGATGATTGA	464
Db	2039	AAGGGAATATTATTAGAAATTTGAAAAAATTTAAACGACAACTAACATTGAAAAACTT	1980
QY	465	AAATACTACTGATTTAACTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	501
Db	1979	AAAAACAATTAATTGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAA	1943

Search completed: July 12, 2005, 11:12:55
Job time : 555.528 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2005, 01:03:44 ; Search time 2468.64 Seconds
(without alignment)
7724.974 Million cell updates/sec

Title: US-10-018-929C-2_COPY_6071_6571
Perfect score: 501
Sequence: 1 caactcatcagttactctct.....cttaaaaaaaaaaaaaaa 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hc3:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	364.4	72.7	566	8 B67281	B67281 T22N9TR TAM
2	357.4	71.3	472	8 B62563	B62563 T22D24TR TA
3	348.4	69.5	639	8 B20434	B20434 T22N9-Sp6 T
4	323.8	64.6	398	1 AU230038	AU230038 AU230038
5	287.2	57.3	661	8 B20425	B20425 T22D24-Sp6
6	188.6	37.5	205	1 AA597930	AA597930 29179 Lam
7	187.8	37.5	401	8 BH212578	BH212578 SALK 0077
8	131	26.1	310	8 B2358017	B2358017 SALK 1317
9	95.6	19.1	454	8 BH746544	BH746544 SALK 0438
10	87	17.4	644	8 B2006642	B2006642 oem85a08
11	78.8	15.7	161	8 BH251820	BH251820 SALK_0121
12	73.2	14.6	635	1 AU238798	AU238798 AU238798
13	63.4	12.7	258	8 BH436380	BH436380 BOHLV76TF
14	50.6	10.1	813	9 BX230908	BX230908 Dario rer
15	47.8	9.5	928	9 CNS00DKY	CNS00DKY Drosophil
16	47.6	9.5	987	9 CNS014PQ	AL104456 Drosophil
17	47.4	9.5	928	9 CNS00DKY	AL071865 Drosophil
18	44.8	8.9	565	7 C0537923	C0537923 tah80e02
19	44.8	8.9	759	9 CNS06QXV	AL411257 T7 end of
20	44.8	8.9	813	9 BX236594	BX236594 Dario rer
21	44.8	8.9	987	9 CNS014PQ	AL104456 Drosophil
22	44.4	8.9	189	1 AU261408	AU261408 AU261408
23	44.4	8.9	843	9 CNS00CS1	AL059666 Drosophil
24	44.4	8.9	1101	9 CNS00EVL	AL069706 Drosophil

C 25	44.2	8.8	528	4 BI399701	BI399701 MI-P-AV1-
C 26	44.2	8.8	872	8 AZ529909	AZ529909 ENTW13TF
C 27	44.2	8.8	945	4 BM358135	BM358135 GA_Ea000
C 28	44	8.8	683	1 AI443491	AI443491 sa32d11.x
C 29	43.8	8.7	271	5 BQ739737	BQ739737 pFESToab4
C 30	43.8	8.7	772	9 AG049468	AG049468 Pan trogl
C 31	43.8	8.7	833	7 CK795655	CK795655 AGENCOURT
C 32	43.4	8.7	241	2 BF612410	BF612410 df07f08.y
C 33	43.4	8.7	791	4 BM162850	BM162850 EST365373
C 34	43.4	8.7	974	9 CNS00ITT	AL05432 Drosophil
C 35	43.2	8.6	1024	9 CNS005BT	AL058090 Drosophil
C 36	43	8.6	422	5 BQ190569	BQ190569 UI-R-D01-
C 37	43	8.6	555	2 BF283441	BF283441 EST448032
C 38	43	8.6	1192	9 CNS03Y36	AL265803 Tetraodon
C 39	42.8	8.5	751	7 CR441500	CR441500 CR441500
C 40	42.8	8.5	908	9 CNS01678	AL106382 Drosophil
C 41	42.8	8.5	2007	3 HSM807665	BX647519 Homo sapi
C 42	42.6	8.5	330	1 AA213260	AA213260 mw85b09.r
C 43	42.4	8.5	732	9 CNS0113J	AL099769 Drosophil
C 44	42.4	8.5	792	8 BH459208	BH459208 BOH9Q92TF
C 45	42.4	8.5	845	8 BH686264	BH686264 BOMP933TR

ALIGNMENTS

RESULT 1
B67281
LOCUS T22N9TR TAMU Arabidopsis thaliana genomic clone T22N9, genomic
DEFINITION survey sequence.
ACCESSION B67281
VERSION B67281.1 GI:2665958
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 566)
AUTHORS Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Adams, M.D. and
Venter, J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL Unpublished (1997)
COMMENT Other GSSs: T22N9TF
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 566.
FEATURES
Location/Qualifiers
1..566
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="T22N9"
/sex="hermaphrodite"
/clone_lib="TAMU"
/note="vector: BelobACII; Site 1: HindIII; Site 2:
HindIII; Produced by Rod Wing"

Query Match 72.7%; Score 364.4; DB 8; Length 566;
Best Local Similarity 98.4%; Pred. No. 3.8e-80;
Matches 368; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 126 CAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACCAACCGCACAG 185
    |||||
Db 1 CAAGCTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACCAACCGCACAG 60

QY 186 AGCTTGAGCAGTGAGTTGCAGAGCAACAATGAAGTGGTTGTCTTTCTGACGAGAGTGA 245
    |||||
Db 61 AGCTTGAGCAGTGAGTTGCAGAGCAACAATGAAGTGGTTGTCTTTCTGACGAGAGTGA 120

QY 246 CCTAAGGAGGAGATGGTTAGGGTCTTAGTATTATTTAGAGAGTTAATAATAGTAT 305
    |||||
Db 121 CCTAAGGAGGAGATGGTTAGGGTCTTAGTATTATTTAGAGAGTTAATAATAGTAT 180

QY 306 ATATATATATGTATAAGTAGGTACCTTAATCTCTGTCGTTAAATCTAATTTAGTGAGTCAG 365
    |||||
Db 181 ATATATATATGTATAAGTAGGTACCTTAATCTCTGTCGTTAAATCTAATTTAGTGAGTCAG 240

QY 366 GAACCGACTCGTTGGCTTAAGGTCTCTCCTTTGAAACGCAACGTTCTACTTTCATGTATA 425
    |||||
Db 241 GAACCGACTCGTTGGCTTAAGGTCTCTCCTTTGAAACGCAACGTTCTACTTTCATGTATA 300

QY 426 TAAATACAGTCTGATCACACACAAATTTGATGATTGAAATACTACTGATTAACTTA 485
    |||||
Db 301 TAAATACAGTCTGATCACACACAAATTTGATGATTGAAATACTACTGATTAACTTT 360

QY 486 AAAAAAAAAAAAAA 499
    |||||
Db 361 ATAGAAACCCAAA 374
```

```
RESULT 2
B62563
LOCUS T22D24TR TAMU Arabidopsis thaliana genomic clone T22D24, genomic
DEFINITION survey sequence.
VERSION B62563.1 GI:2629325
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 472)
Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
Venter,J.C.
A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
Unpublished (1997)
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 472.
FEATURES
    source
        1..472
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clone="T22D24"
            /sex="hermaphrodite"
            /clone_lib="TAMU"
            /note="vector: BelobACII; Site_1: HindIII; Site_2:
            HindIII; Produced by Rod Wing"
```

```
Query Match 71.3%; Score 357.4; DB 8; Length 472;
Best Local Similarity 98.4%; Pred. No. 2.1e-78;
Matches 361; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 133 TGACATATTTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACCAACCGCACAGCTTGA 192
    |||||
Db 1 TGACATATTTCAGCTGTGTCAATTCAGCAGCAGCAAGCAACCAACCGCACAGCTTGA 60

QY 193 GCAGTGGATTGCAGAGCAACAATGAAGTGGTTGTCTTTCTGACGAGAGTGAACCTAAGA 252
    |||||
Db 61 GCAGTGGATTGCAGAGCAACAATGAAGTGGTTGTCTTTCTGACGAGAGTGAACCTAAGA 120

QY 253 GGAGAGATGGTTAGGGTCTTAGTATTATTTAGAGAGTTAATAATAGTATATATA 312
    |||||
Db 121 GGAGAGATGGTTAGGGTCTTAGTATTATTTAGAGAGTTAATAATAGTATATATA 180

QY 313 TAATGATTAAGTAGGTACCTTAATCTCTGTCGTTAAATCTAATTTAGTGAGTCAGAACCGA 372
    |||||
Db 181 TAATGATTAAGTAGGTACCTTAATCTCTGTCGTTAAATCTAATTTAGTGAGTCAGAACCGA 240

QY 373 CTGCTTGGCTTAAGGTCTCTCCTTTTGAACGCAACGTTCTACCTTTTCATGTATATAAATAC 432
    |||||
Db 241 CTGCTTGGCTTAAGGTCTCTCCTTTTGAACGCAACGTTCTACCTTTTCATGTATATAAATAC 300

QY 433 AGTCTGATCACACACACAAATTTGATGATTGAAAAATCTACTGATTAACTTAAAAAAA 492
    |||||
Db 301 AGTCTGATCACACACAAATTTGATGATTGAAAAATCTACTGATTAACTTAAAAAAA 360

QY 493 AAAAAAA 499
    |||||
Db 361 ACCCAAA 367

RESULT 3
B20434
LOCUS T22N9-Sp6 TAMU Arabidopsis thaliana genomic clone T22N9, genomic
DEFINITION survey sequence.
VERSION B20434.1 GI:2395488
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 639)
Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
BAC End Sequences at ATGC
Unpublished (1997)
Other_GSSs: T22N9-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: Jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 95
High quality sequence stop: 382.
FEATURES
    source
        1..639
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clone="T22N9"
            /sex="hermaphrodite"
            /clone_lib="TAMU"
```



```
High quality sequence stop: 300.
FEATURES
  source
    1..661
      Location/Qualifiers
        /organism="Arabidopsis thaliana"
        /mol_type="genomic DNA"
        /ecotype="Columbia"
        /db_xref="taxon:3702"
        /clone="T22D24"
        /sex="hermaphrodite"
        /clone_lib="TAMU"
        /notes="vector: BelobACII; Site 1: HindIII; Site 2: HindIII; Produced by Rod Wing"
ORIGIN
  Query Match      57.38; Score 287.2; DB 8; Length 661;
  Best Local Similarity 85.4%; Pred. No. 7.1e-61;
  Matches 316; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 124 CTCAGCTTTTGACATATTTCAGCTGTGTCAATTTCAGCAGCAGCAAGAACCAACCGCAAC 183
DB 56 CCNNGCTCCGAATATNCGCTGTGTCTTTCATCTTCANCANCAAGAACCAACCGCAAC 115
QY 184 AGAGCTTGAGCAGTGGATTCGAGAGCAACAATGAAGTGGTTTGTCTTTCTGACGACGAT 243
DB 116 TGAGCTTGAGCAGTGCATCGCACCAACAATGAATGGTTTGTCTTTCTGACGACTAGT 175
QY 244 GACCTAAGAGAGAGATGGTTAGGGTCTTAGTATTGATTTTTCAGAGATTAATAGT 303
DB 176 GACCTAAGAGAGAGATGGTCAAGGCTCTCACTCAATGATTTTTCAGAGATTAATAGT 235
QY 304 ATATATATATATATATGAAGTAGGTCTCAATATCTCTGCTGTTAATCTAAATTAGTAGTC 363
DB 236 CTCTATATCTTGTCTCAGTCGGTCCCTAATCTCTGCTGTCATCTAATTTAGTAGTC 295
QY 364 AGGAACCGACTCGTTGGCTCTAAGGTCTCTCCCTTTTGAACCGAAGCTTCTACTTTTCATGTA 423
DB 296 AGGAACCGACTCGTTGGCTCTAAGGTCTCTCCCTTTTGAACCGAAGCTTCTACTTTCCATGTA 355
QY 424 TATTAATACAGTCTGATCACACACACAAATGATGATTTGAATATCTACTGATTTAACT 483
DB 356 TATCAATCCAGTCTGATCCCAACCAACGATCGATCGAACAATCTACTGACTTTACT 415
QY 484 TAAAAAAAAA 493
DB 416 TCATCGAACA 425
RESULT 6
AA597930 205 bp mRNA linear EST 19-SEP-1997
LOCUS 23179 Lambda-PRL2 Arabidopsis thaliana cDNA clone 247N2T7, mRNA
DEFINITION
  sequence.
ACCESSION
  AA597930
VERSION
  AA597930.1 GI:2413353
KEYWORDS
  EST.
SOURCE
  Arabidopsis thaliana (thale cress)
  ORGANISM
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 1 to 205)
REFERENCE
  AUTHORS
    Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,
    McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M.,
    Retzel, E., and Somerville, C.
  TITLE
    Genes galore: a summary of methods for accessing results from
    large-scale partial sequencing of anonymous Arabidopsis cDNA clones
    Plant Physiol. 106, 1241-1255 (1994)
  JOURNAL
    95148729
  MEDLINE
    7846151
  PUBMED
    Contact: Thomas Newman
    MSU-DOE Plant Research Laboratory
    Michigan State University
    MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, Mi
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcm@ibm.cl.msu.edu
Seq primer: T7 dye primer.
FEATURES
  Location/Qualifiers
    source
      1..205
        /organism="Arabidopsis thaliana"
        /mol_type="mRNA"
        /ecotype="Columbia"
        /db_xref="taxon:3702"
        /clone="247N2T7"
        /clone_lib="Lambda-PRL2"
        /notes="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not;
        Lambda PRL2 is a cDNA library derived from equal_
        quantities of 4 pools of mRNA. The mRNA sources were 1) 7
        day germinated etiolated seedlings; 2) tissue culture
        grown roots; 3) staged plants half with 24 hour light
        cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
        same plants as 3 but aerial tissue (stems, flowers and
        siliques. The vector is BRL's lambda Zip-Lox. The cDNA
        inserts were directionally cloned with Sal-Not arms using
        oligo dt primed cDNA. "
ORIGIN
  Query Match      37.6%; Score 188.6; DB 1; Length 205;
  Best Local Similarity 94.6%; Pred. No. 2.4e-36;
  Matches 191; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 292 GTTAAATAAGTATATATATATATATATATATATATATATATATATATATATATATAT 351
DB 1 GTTAAATAAGTATATATATATATATATATATATATATATATATATATATATATATAT 60
QY 352 ATTTAGTGTGAGGACCGACTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 411
DB 61 ATTNAGTGTGAGGACCGACTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 120
QY 412 TACTTTCATGTATATAAATACAGTCTGATCACAACACAAATTTGATTTGAAAAATACT 471
DB 121 TACTTTCATGTATATAAATACAGTCTGATCACAACACAAATTTGATTTGAAAAATACN 180
QY 472 ACTGATTTAACTTTAAAAAAA 493
DB 181 ACTGATTTAACTTTTATAGAAA 202
RESULT 7
BH212578/c
LOCUS BH212578
DEFINITION
  SALK_007765 Arabidopsis thaliana TDNA insertion lines Arabidopsis
  thaliana genomic clone SALK_007765, genomic survey sequence.
ACCESSION
  BH212578
VERSION
  BH212578.1 GI:16393811
KEYWORDS
  GSS.
SOURCE
  Arabidopsis thaliana (thale cress)
  ORGANISM
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 1 to 401)
REFERENCE
  AUTHORS
    Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
    Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
    Shinn, P., Zimmerman, J., and Ecker, J.R.
  TITLE
    A Sequence-Indexed Library of Insertion Mutations in the
    Arabidopsis Genome
  JOURNAL
    Unpublished (2001)
  COMMENT
    Contact: Joseph R. Ecker
    Salk Institute Genomic Analysis Laboratory (SIGNAL)
    The Salk Institute for Biological Studies
    10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
    Tel: 858 453 4100 x1752
    Fax: 858 558 6379
    Email: ecker@salk.edu
```

This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 3' end of At1g08060.
Class: TDNA tagged.

FEATURES
source
1. 401
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK 007765"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 37.5%; Score 187.8; DB 8; Length 401;
Best Local Similarity 92.1%; Pred. No. 4.2e-36;
Matches 198; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 111 TCGTCAGTGCCTCTCAAGCTTTGCACATATTCAGCTGTGTCATTCAGCAGCACAAGAA 170
Db 220 TCGACGCCCGGGCTGCAGCTTTGCACATATTCAGCTGTGTCATTCAGCAGCACAAGAA 161
Qy 171 CAACAACCGCAACAGAGCTTCGAGCAGTGGATTGAGAGCAACAATGAAGTGTGCTT 230
Db 160 CAACAACCGCAACAGAGCTTCGAGCAGTGGATTGAGAGCAACAATGAAGTGTGCTT 101
Qy 231 TCTGACGACGAGTGCACCTAAGAGAGAGATGTTAGGTCCTTACTTATTCATTTTATAGAG 290
Db 100 TCTGACGACGAGTGCACCTAAGAGAGAGATGTTAGGTCCTTACTTATTCATTTTATAGAG 41
Qy 291 AGTTAATAATAGTAT 325
Db 40 AGTTAATAATAGTAT 6

RESULT 8
BZ358017
LOCUS
DEFINITION
BZ358017.1 GI:24950134 310 bp DNA linear GSS 14-NOV-2002
SALK_131757.53.10.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_131757.53.10.x, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
The Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

TITLE
JOURNAL
COMMENT
This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g08060.
Class: TDNA tagged.

FEATURES
Location/Qualifiers

1. 310
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_131757.53.10.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 26.1%; Score 131; DB 8; Length 310;
Best Local Similarity 100.0%; Pred. No. 6e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATCCATCAGTTACTCTCTGCAACAATCCAGGTCTGCGTTCTCTGCAACACCTAA 60
Db 180 CAATCCATCAGTTACTCTCTGCAACAATCCAGGTCTGCGTTCTCTGCAACACCTAA 239
Qy 61 ACTCATATAGACCATCTCTTCACTCCGTCGCCAGAGTACTCCAACTCGTCAGTGC 120
Db 240 ACTCATATAGACCATCTCTTCACTCCGTCGCCAGAGTACTCCAACTCGTCAGTGC 299
Qy 121 CTCCTCAAGCT 131
Db 300 CTCCTCAAGCT 310

RESULT 9
BH746544/c
LOCUS
DEFINITION
BH746544.1 GI:18959659 454 bp DNA linear GSS 27-FEB-2002
SALK_043884.28.75.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_043884.28.75.x, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
The Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

TITLE
JOURNAL
COMMENT
This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g08060.
Class: TDNA tagged.

FEATURES
source
1. 454
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="SALK_043884.28.75.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tadna_protocols.html

	Query Match	19.1%	Score 95.6;	DB 8;	Length 454;
	Best Local Similarity	75.3%;	Pred. No. 4.2e-13;		
	Matches 119;	Conservative 0;	Mismatches 39;	Indels 0;	Gaps 0;
Qy	62	CTCATATAGACCATCCTCTTCAACTCCGGTCGCCACAGCTACTCCAACCTCGTCAGTGCC	121		
Db	168	CGCATGAAGATGAGTCTTTTTTTACGACTCACTATTAGGGCACGCGTGGTCGACGGCCCG	109		
Qy	122	TCCTCAAGCTTTGCACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGAAACAAACCCGCA	181		
Db	108	GTTTGCAGCTTTGCATATTCAGCTGTGTCAATTCAGCAGCAGCAAGAAACAAACCCGCA	49		
Qy	182	ACAGAGCTTCAGCAGTGGATTGTCAGAGCAACAATGAAG	219		
Db	48	ACAGAGCTTCAGCAGTGGATTGTCAGAGCAACAATGAAG	11		

RESULT 10	BZ006642/c
LOCUS	BZ006642
DEFINITION	oem85a08.g1 B.oleracea002 Brassica oleracea genomic, genomic survey sequence.
ACCESSION	BZ006642
VERSION	BZ006642.1
KEYWORDS	GI:23554900
SOURCE	GSS.
ORGANISM	Brassica oleracea
	Brassica oleracea
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE	1 (bases 1 to 644)
AUTHORS	Delehaenty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE	Whole genome shotgun reads from Brassica oleracea
JOURNAL	Unpublished (2002)
COMMENT	Contact: Richard K. Wilson

```

FEATURES
source
n-gn quality sequence scop: 351.
Location/Qualifiers
1. .644
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

```

ORIGIN	Sequencing Center.
Query Match	17.4%; Score 87; DB 8; Length 644;
Best Local Similarity	64.1%; Pred. No. 6.2e-11;
Matches 184; Conservative	0; Mismatches 90; Indels 13; Gaps 3;
Qy	75 TCCTCTTCAACTCCCGTCGCC--ACAGCTACTCCAACTCGTCAGTGCCTCTCTCAAGCT 131
Db	433 TCCTCTTCAGTTCCCATCCCAAGAACTACTTCAACCTCATCAGGTGTTCTCTCAAGCT 374

[illegible]

RESULT 11	
BH251820/c	
LOCUS	BH251820 161 bp DNA linear GSS 28-NOV-2001
DEFINITION	SALK_012181 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_012181, genomic survey sequence.
ACCESSION	BH251820
VERSION	BH251820
KEYWORDS	BH251820.1 GI:17138798
SOURCE	GSS.
ORGANISM	Arabidopsis thaliana (thale cress)
REFERENCE	Arabidopsis thaliana
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 161)
TITLE	Alonso J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
JOURNAL	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
COMMENT	Unpublished (2001) Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g08060. Class: TDNA tagged.

FEATURES	source
Cross. and tagged	Location/Qualifiers
	1. .161
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/ecotype="Col-0"
	/db_xref="taxon:3702"
	/clone="SALK_012181"
	/clone_lib="Arabidopsis thaliana TDNA insertion lines"
	/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html "

	Query Match	15.7%	Score 78.8;	DB 8;	Length 161;
	Best Local Similarity	79.1%;	Pred. No. 5.8e-09;		
	Matches 106;	Conservative 0;	Mismatches 27;	Indels 1;	Gaps 1;
Qy	111	TCGTCAGTGCTCTCAAGCTTTTGACATATTCAGCTGTGTCAATTCAGCAGCAGCAAGAA	170		
Db	134	TCGACGCCCGGGGTGCAGCTTTTGACATATTCAGCTGTGTCAATTCATCAGCAGCAGAT	75		
Qy	171	CAACAACCGCAACAGAGCTTTGACAGCTGGATTGCAGAGCAAC-AATGAAGTGTTTCTCT	229		

Db 74 CGGCCACGGCAGAGCGGTGACCACTGGCGATGAGAGCTACAAAGGATCGGATTGCTCT 15

Qy 230 TTTCGACGACGAGT 243
|||||

Db 14 TTCTGCCGACAGT 1

RESULT 12
AU238798 635 bp mRNA linear EST 01-APR-2002
LOCUS AU238798 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-14-009 5',
DEFINITION mRNA sequence.
ACCESSION AU238798
VERSION AU238798.1 GI:19877967
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids;
1 (bases 1 to 635)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A.,
Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source Location/Qualifiers
1..635
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL19-14-009"
/tissue_types="mixture of silique and flower"
/lab_host="DH10B"
/clone_lib="RAFL19"
/note="Site_1: BamHI; Site_2: SalI; Subtraction Library"

ORIGIN
Query Match 14.6%; Score 73.2; DB 1; Length 635;
Best Local Similarity 88.5%; Pred. No. 1.7e-07;
Matches 123; Conservative 0; Mismatches 11; Indels 5; Gaps 4;
Qy 1 CAACTCCATCAGTTACTCTCGCAAAATCCA-GGTCTGGTCTCTCTGCACACACCTA 59
Db 497 CAACTCCATCAGTTACTCTCTGCAAAATCCAAGGCTCTGGTCTCTCTGCACACACCTA 556
Qy 60 AA-CTCATATAGACCATCTCTTC-AACTCCCGTCGCGCAGAGTACTCCAA--CCTCGTC 115
Db 557 AACCTCATATAGACCATCTCTTTCAAACTCCGTCGCGCAGAGTACTCTCAAACTCGTCA 616
Qy 116 AGTCCCTCTCAAGCTTTG 134
Db 617 AGTCCCTCTCAAGCTTTG 635

RESULT 13
BH436380 258 bp DNA linear GSS 12-DEC-2001
LOCUS BH436380

DEFINITION BOHLV76TF BOHL Brassica oleracea genomic clone BOHLV76, genomic
survey sequence.
ACCESSION BH436380
VERSION BH436380.1 GI:17622101
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 258)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOHLV76TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..258
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="BOHLV76"
/notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match 12.7%; Score 63.4; DB 8; Length 258;
Best Local Similarity 63.0%; Pred. No. 4.3e-05;
Matches 114; Conservative 0; Mismatches 66; Indels 1; Gaps 1;
Qy 140 TTCAGCTGTCTCAATTTCAGCAGCAGCAAGAACCAACCGCACAGAGCTTGACGAGTGG 199
Db 36 TTTACCATATTCAAATGCGTTTATTTCAGCAGCAGCAAGAACCAACCGAGACTTGGGAGTGG 95
Qy 200 ATTGCAGAGCAACAATGAAGTGGTTTGTCTTCTTCGACGAGTGACCTAAGAGAGAGA 259
Db 96 ATTGCAGAGGAGCAATGATGTGTCTCTCTCTCTGATGACGAGTGATCTAA-TCTTCTTA 154
Qy 260 TGGTTAGGCTTCTAGTTATTGATTTTGTAGAGAGTTAATAATAGTATATATATATAT 319
Db 155 TGTTAGCGTTTAATAGTTTGGTCTAGAGTTGTATATAAAAAAAGTTAAGTTACCTATAAT 214
Qy 320 A 320
Db 215 A 215

RESULT 14
BX230908/c 813 bp DNA linear GSS 29-JAN-2003
LOCUS BX230908
DEFINITION Danio rerio genomic clone DKEY-53M12, genomic survey sequence.
ACCESSION BX230908
VERSION BX230908.1 GI:28065058
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 813)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2005, 13:13:50 ; Search time 63 Seconds
(without alignments)
3812.352 Million cell updates/sec

Title: US-10-018-929c-3_COPY_650_1270

Perfect score: 3216

Sequence: 1 RADAFILFGSLNPSHDVKH.....SGNNKHFGSSNISONAPDV 621

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3216	100.0	2001	4 AAB20062	Aab20062 Arabidops
2	3207	99.7	2001	4 AAB20063	Aab20063 Arabidops
3	163.5	5.1	1465	8 ADN22447	Adn22447 Bacterial
4	162.5	5.1	842	8 ADQ66710	Adq66710 Novel hum
5	162.5	5.1	1739	6 AAE36105	Aae36105 Human chr
6	162.5	5.1	1739	6 AAE36106	Aae36106 Human chr
7	162.5	5.1	1739	8 ADK60208	Adk60208 Angiogene
8	162.5	5.1	1739	8 ADL12347	Adl12347 Human ste
9	162.5	5.1	1739	8 ADK60509	Adk60509 Angiogene
10	162.5	5.1	1739	8 ADP73132	Adp73132 Angiogene
11	158	4.9	1015	8 ADN20557	Adn20557 Bacterial
12	156	4.9	954	4 AAU14615	Aau14615 Novel bon
13	155.5	4.8	2228	7 ABR61599	Abr61599 Human gol
14	155.5	4.8	2230	6 ABU07445	Abu07445 Protein d
15	155.5	4.8	2230	6 ABR61600	Abr61600 Human gol
16	155.5	4.8	2250	7 ABR61601	Abr61601 Human gol
17	155.5	4.8	2252	7 ABR61602	Abr61602 Human gol
18	149.5	4.6	823	4 ABUS3074	Abu53074 Intracell
19	149.5	4.6	875	4 ABUS3078	Abu53078 Intracell
20	148	4.6	1455	7 ADE80783	Ade80783 Microsate
21	147.5	4.6	1959	5 ABJ10604	Abj10604 Human nov
22	147	4.6	962	3 AAB18211	Aab18211 Plasmodiu
23	147	4.6	1948	4 ABG21233	Abg21233 Novel hum
24	146	4.5	1738	8 ABM83885	Abm83885 Human dia
25	146	4.5	1881	5 ABP73809	Abp73809 Candida a

26	145.5	4.5	1647	5 ABP73735	Abp73735 Candida a
27	145.5	4.5	1931	4 ABB61012	Abb61012 Drosophil
28	145	4.5	1960	4 AAM78854	Aam78854 Human pro
29	145	4.5	1960	7 ADB70370	Adb70370 PAC 6802
30	145	4.5	1960	7 ADD47946	Ad47946 Human pro
31	145	4.5	1960	8 ADJ37149	Adj37149 Human mal
32	145	4.5	1960	8 ABM82345	Abm82345 Tumour-as
33	145	4.5	1960	8 ADN60283	Adn60283 Human non
34	145	4.5	2143	4 ABG01716	Abg01716 Novel hum
35	145	4.5	3187	7 ADE56031	Ade56031 Rat Prote
36	145	4.5	3187	7 ADE56035	Ade56035 Rat Prote
37	145	4.5	3225	7 ADJ68448	Adj68448 Human hea
38	143.5	4.5	1427	2 AAR10534	Aar10534 Human 160
39	143	4.4	1130	8 ADN24098	Adn24098 Bacterial
40	143	4.4	1379	7 ADB79765	Adb79765 Rat Rho k
41	142.5	4.4	1963	4 AAM79838	Aam79838 Human pro
42	142	4.4	951	4 ABUS3070	Abu53070 Intracell
43	142	4.4	953	4 ABUS3071	Abu53071 Intracell
44	141.5	4.4	1392	2 AAY06999	Aay06999 Restin pr
45	140.5	4.4	3210	6 ABU07438	Abu07438 Protein d

ALIGNMENTS

RESULT 1

AAB20062
ID AAB20062 standard; protein; 2001 AA.

XX AC AAB20062;

XX AC AAB20062;

DT 04-NOV-2004 (revised)

DT 23-APR-2001 (first entry)

XX XX

DE Arabidopsis thaliana silencing gene-encoded protein.

XX Gene silencing; silencing gene; MOM.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT Region 177..350

FT /note= "internal repeat region"

FT Peptide 362..367

FT /note= "nuclear localization motif"

FT Binding-site 460..467

FT /note= "ATP/GTP-binding motif"

FT Region 479..719

FT /note= "region of homology to ATPase/helicase family

FT Misc-difference 705

FT /note= "Lys in ecotype Columbia"

FT Peptide 832..838

FT /note= "nuclear localization motif"

FT Peptide 858..862

FT /note= "nuclear localization motif"

FT Domain 995..1015

FT /note= "predicted membrane-spanning domain"

FT Misc-difference 1219

FT /note= "Asp in ecotype Columbia"

FT Region 1462..1672

FT /note= "internal repeat region"

FT Region 1848..1894

FT /note= "internal repeat region"

FT Domain 1899..1941

FT /note= "actin-binding domain"

XX WC200100801-A2.

XX 04-JAN-2001.

XX 21-JUN-2000; 2000WO-EP005761.

XX

```
PR 23-JUN-1999; 99GB-00014623.
XX (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
DR N-PSDB; AAA89353, AAA89354.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
PT
XX
XX Claim 1; Page 32-39; 48pp; English.
XX
XX The present sequence is that of Arabidopsis thaliana protein MOM, which
CC controls gene silencing, particularly gene silencing. It is encoded by a
CC novel silencer gene (see AAA89353) of A. thaliana ecotype Zurich. The
CC protein exhibits partial similarity with ATPase/helicase proteins of the
CC SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
CC compared with the protein (see AAB20063) encoded by the silencer gene of
CC wild-type A. thaliana ecotype Columbia. Gene silencing is useful as a
CC molecular tool for regulating gene expression
CC
CC Revised record issued on 04-NOV-2004 : Correction to feature table key
XX
XX Sequence 2001 AA;
SQ
Query Match 100.0%; Score 3216; DB 4; Length 2001;
Best Local Similarity 100.0%; Pred. NO. 4e-267;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RADAFILFGSSLPNSHDVGHVEKIKIESCSERKIFRLYSVCTVEEKALILARQNRONK 60
DB 650 RADAFILFGSSLPNSHDVGHVEKIKIESCSERKIFRLYSVCTVEEKALILARQNRQK 709
QY 61 AVENLNRSLTTHALLMWGASYLFKLDHFHSSETPDSCVSPESQSIMDGVIEHFSLSISKG 120
DB 710 AVENLNRSLTTHALLMWGASYLFKLDHFHSSETPDSCVSPESQSIMDGVIEHFSLSISKG 769
QY 121 GEENEVKLCILLPAKHAQGYTSSDSTLFGEDHIKLSDESPNIFWSKLGKGNPMWKYPS 180
DB 770 GEENEVKLCILLPAKHAQGYTSSDSTLFGEDHIKLSDESPNIFWSKLGKGNPMWKYPS 829
QY 181 DTPQRNRKRVQYFEGSEASPKTGGGNNAKRRKASDDVTPRVTDPVDDDERKASGKH 240
DB 830 DTPQRNRKRVQYFEGSEASPKTGGGNNAKRRKASDDVTPRVTDPVDDDERKASGKH 889
QY 241 MGALESPPKVTITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 300
DB 890 MGALESPPKVTITLQSSCKSSGTDGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 949
QY 301 QRLRLTVLPKMAKLCQVLHLSDACTSMVGNFLEYIENHRIYEEPPATTFQAFQIALSWI 360
DB 950 QRLRLTVLPKMAKLCQVLHLSDACTSMVGNFLEYIENHRIYEEPPATTFQAFQIALSWI 1009
QY 361 AALLVQIILSHKESLVRANSELAFKCSRVEVDYIYTSLSCKMSLFLEHTQGLQFDCGFTN 420
DB 1010 AALLVQIILSHKESLVRANSELAFKCSRVEVDYIYTSLSCKMSLFLEHTQGLQFDCGFTN 1069
QY 421 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCCHYSTATRDIEKTIS 480
DB 1070 SKQSVVSTKLVNESLSGATVRDEKINTKSMRNSSEDEECMTKRCCHYSTATRDIEKTIS 1129
QY 481 GIKKYKQVQKLVQBEHEKMKELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 540
DB 1130 GIKKYKQVQKLVQBEHEKMKELLNNYADKKQKLETSKSVEAAVIRITCSRTSTQVGD 1189
QY 541 KLADHNYERKFDRIKSEKNECLSLQEMHVEVAKKLAEDACWNRITKSWAAKLKVCVPI 600
DB 1190 KLADHNYERKFDRIKSEKNECLSLQEMHVEVAKKLAEDACWNRITKSWAAKLKVCVPI 1249
QY 601 QSGNNKHFGSSNISQNA PDV 621
DB 1250 QSGNNKHFGSSNISQNA PDV 1270
RESULT 2
AAB20063
ID AAB20063 standard; protein; 2001 AA.
XX
XX AAB20063;
XX
XX 23-APR-2001 (first entry)
XX
XX Arabidopsis thaliana silencing gene-encoded protein.
XX
XX Gene silencing; silencing gene; MOM.
XX
XX Arabidopsis thaliana.
XX Key Location/Qualifiers
XX Region 177..350
XX /note= "internal repeat region"
XX Peptide 362..367
XX /note= "nuclear localization motif"
XX Binding-site 460..467
XX /note= "ATP/GTP-binding motif"
XX Region 479..719
XX /note= "region of homology to ATPase/helicase family
XX SWI2/SNF proteins"
XX Misc-difference 705
XX /note= "Met in ecotype Zurich"
XX Peptide 832..838
XX /note= "nuclear localization motif"
XX Peptide 858..862
XX /note= "nuclear localization motif"
XX Domain 995..1015
XX /note= "predicted membrane-spanning domain"
XX Misc-difference 1219
XX /note= "Glu in ecotype Zurich"
XX Region 1462..1672
XX /note= "internal repeat region"
XX Region 1848..1894
XX /note= "internal repeat region"
XX Domain 1899..1941
XX /note= "actin-binding domain"
XX
XX WO200100801-A2.
XX
XX 04-JAN-2001.
XX
XX 21-JUN-2000; 2000WO-EP005761.
XX
XX 23-JUN-1999; 99GB-00014623.
XX
XX (NOVS ) NOVARTIS AG.
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
XX WPI; 2001-137952/14.
DR N-PSDB; AAA89385.
XX
XX Novel gene encoding a protein that controls gene silencing, in particular
PT silencing of plant genes.
PT
XX
XX Claim 8; Page; 48pp; English.
XX
XX The present sequence is that of Arabidopsis thaliana protein MOM, which
CC controls gene silencing, particularly gene silencing. It is encoded by a
CC novel silencer gene (see AAA89385) of A. thaliana ecotype Columbia. The
CC protein exhibits partial similarity with ATPase/helicase proteins of the
CC SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219
CC compared with the protein (see AAB20062) encoded by the silencer gene of
CC Arabidopsis thaliana ecotype Zurich. Gene silencing is useful as a
```


CC molecular tool for regulating gene expression. Note: the present sequence
CC is not shown in the specification but is derived from the A. thaliana
CC mutant silencer gene protein sequence given in AAB20062
XX
SQ Sequence 2001 AA;

Query Match 99.78; Score 3207; DB 4; Length 2001;
Best Local Similarity 99.78; Pred. No. 2.4e-266;
Matches 619; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RADAFILFGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQNK 60
DB 650 RADAFILFGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQNK 709
QY 61 AVENLNRLTHALLMWGASYLFDKLDHFSSETPDGSGVSEFQSIQMDGVIHEFSGILSSKG 120
DB 710 AVENLNRLTHALLMWGASYLFDKLDHFSSETPDGSGVSEFQSIQMDGVIHEFSGILSSKG 769
QY 121 GEENEVKLCILLEAKHAQGYSSDSTLFGEDHFKLSDEESPNFWSKLGCKPMWKYPS 180
DB 770 GEENEVKLCILLEAKHAQGYSSDSTLFGEDHFKLSDEESPNFWSKLGCKPMWKYPS 829
QY 181 DTPQNRKRYQYFEGSEASPKTGGNAKKRKKASDDVTPRVTDPPVDDDERKASKDH 240
DB 830 DTPQNRKRYQYFEGSEASPKTGGNAKKRKKASDDVTPRVTDPPVDDDERKASKDH 889
QY 241 MGALESKVITLQSSCKSSGDTGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 300
DB 890 MGALESKVITLQSSCKSSGDTGLDGNDAFLGYSMGSHISGIPEDMLASQDWGKIPDES 949
QY 301 QRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVNIENRIYEEPATTTQAFQIALSWI 360
DB 950 QRLHTVLKPKMAKLCQVHLSDACTSMVGNFLEYVNIENRIYEEPATTTQAFQIALSWI 1009
QY 361 AALLVKQILSHKSLVRANSELAFKSRVVDYTYSTLSCWKSFLFLEHTQLOPDCFGTN 420
DB 1010 AALLVKQILSHKSLVRANSELAFKSRVVDYTYSTLSCWKSFLFLEHTQLOPDCFGTN 1069
QY 421 SKQSVSTKLVNESLSGATVRDEKINTKSMRNSSEDECKTCRCSHYSTATRDIKTIS 480
DB 1070 SKQSVSTKLVNESLSGATVRDEKINTKSMRNSSEDECKTCRCSHYSTATRDIKTIS 1129
QY 481 GIKKKYKQVKLVQEHKEKMLLNMYADKKQLETSKSVEAAVIRITSRTSTQVGD 540
DB 1130 GIKKKYKQVKLVQEHKEKMLLNMYADKKQLETSKSVEAAVIRITSRTSTQVGD 1189
QY 541 KLLDHNVERKFDKTSKNEKLSLEOMHEVAKKLAEDACWNRITKSWAAKLKVCVPI 600
DB 1190 KLLDHNVERKFDKTSKNEKLSLEOMHEVAKKLAEDACWNRITKSWAAKLKVCVPI 1249
QY 601 QSGNKHFGSSNISQNPDV 621
DB 1250 QSGNKHFGSSNISQNPDV 1270

RESULT 3
ADN22447
ID ADN22447 standard; protein; 1465 AA.

XX AC ADN22447;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polypeptide #5100.

XX KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polypeptide.

OS Bacteria.

XX PN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WI; 2004-061375/06.

XX DR New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.

XX PS Claim 1; SEQ ID NO 5100; 122pp; English.

XX CC The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 1465 AA;

Query Match 5.1%; Score 163.5; DB 8; Length 1465;
Best Local Similarity 17.6%; Pred. No. 0.00045;
Matches 126; Conservative 95; Mismatches 236; Indels 257; Gaps 27;

QY 2 ADAFILFGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKALILARQNMQNK 61

DB 799 ADTVILFSDWNPNQDLQANSRAHRIQTQTVNIYRLVTKGSVEE-----EI 845

QY 62 VENLNRLTHALLMWGASYLFDKLDH-----FHSSETPDGSGVSEFQSIQMDGV 108

DB 846 VERAKRKLV-----LDHLVIQMDTTGKTVLSKNATASGSPFQKLSAI 891

QY 109 IHFESSI--LSSKGBENEV-----KCLLLEAKHAQGYSSDSTLFGEDHFKLSDEES 160

DB 892 L-RFGAVELPKKEGEPEFEDIDRILMGAETREAEVEEYKNEKLLSSPKYANFAIDEE 950

QY 161 PNIF-----WSKLLGGKPMWKYPSDTPQNRKRV-----Q 191

DB 951 KDIATAADEWAAII-----FEEDRNRILEEERKELAEWNLAPRQRKQIPQ 997

KW cytostatic.
XX Homo sapiens.
XX WO200298899-A2.
XX PD 12-DEC-2002.
XX 03-JUN-2002; 2002WO-US017466.
XX 05-JUN-2001; 2001US-0296076P.
XX PR 10-OCT-2001; 2001US-0328605P.
XX PR 22-OCT-2001; 2001US-0338733P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX PI Lioubin MN;
XX DR WPI; 2003-156840/15.
XX PT Identifying a candidate p53 pathway-modulating agent as therapeutic
XX PT targets for disorders related to defective p53 function e.g. cancer by
XX FT contacting an assay system having purified CHD polypeptide or nucleic
XX FT acid, with a test agent.
XX PS Claim 13; Page 185-192; 278pp; English.
XX CC The present invention relates to a method for identifying candidate p53
XX CC pathway modulating agents. The method involves contacting an assay system
XX CC comprising purified chromatin organisation modifier (chromo) domain
XX CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
XX CC active fragments or derivatives, with a test agent under conditions
XX CC where, but for the presence of the test agent, the system provides a
XX CC reference activity. The methods are useful for identifying modulators of
XX CC the p53 pathway as therapeutic targets for disorders associated with
XX CC defective p53 function, such as angiogenic disorders, apoptotic disorders
XX CC or cell proliferative disorders, e.g. cancer. The modulators are useful
XX CC as research reagents, diagnostics and therapeutics. The invention is also
XX CC useful in gene therapy. The present sequence is human chromodomain
XX CC helicase DNA binding protein.
XX SQ Sequence 1739 AA;
Query Match 5.1%; Score 162.5; DB 6; Length 1739;
Best Local Similarity 18.9%; Pred No. 0.00071;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
2 ADAPILGSSLNPSHDVHKVHKIKIESCSERTKIFRLYSVCTVEKA-----L 49
879 ADTVIFDSDWNPQDLQAARAHRIQKQVNIYRLVTKGTVEEIIERAKKQWLDHL 938
50 ILARQNRQKAVEN-----LNRSLTHALLMWGASYLFDKLDHHSSETPDGVSF 100
939 VIQRMDTTGRTILENNSGRSNPNFNEBELTAILKFGAEDLFKELE-----GSES 988
101 EQSIMDGVIEHFFSITLSSKGEENEVKLLLEAKHAQGTYSSTLFGEDHIKLSDRES 160
989 EPQEMD-----IDBILRLAETRENEVSTSTADELL-SQKVFANFATMEDEEL-----EER 1038
161 PNIFWSKLLGGKPNMWPYSPDTPQNRKRVQYFE-----GSEASPKTG 203
1039 PHKDWEII-----PEQRKKEVEERQKELEBIYMLPRISSTKKAQTND 1085
204 DGGNAKRRKKASDDVTPDPRVDDDERKASGDHMGALSKVITLQSCSKSSGTDG 263
1086 DSDTESKQAKQRSASESETED---SDDDKK-----PKRRGRPRSVKDLVVEG 1130
264 TLDGN-----DAFGLYNGSHISGIPEDML--ASQDNGKIPDESORLHTVLKPKMAKLC 316
1131 FTDAEIRRFKAYKKF-----GLFLERLECLARD-AELVDKSVADL-----KRLG 1174

QY 317 QVHLSDACTSMVGNFLEYVIEHRIYEEPPATTPQAFQIALSWIAALLVQLLSHESLV 376
DB 1175 ELIH--NSCVSAMQEBEQLKENASEGKGPGR-RGPTIKISGV-QVNVKSIIOHEBPE 1230
QY 377 RANSEL-----AFKCSRVEDVDIYSTLSCKMSLFLHETGLODFDCFGTNSK 422
DB 1231 MLHKSIPVDPEKKKYCLTCRVKAHPDVEGVEDDSRLLLGIYEHGYG-NWELIKTDP 1289
QY 423 QSVVSTKLNVESLSGATVRDEKINTKSMRNSSE-----DEECMTKR 464
DB 1290 -----LKITDKILPVE--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGEEAKLKR 1342
QY 465 -----CSHYSTATRDTEKTSIGIKYKVKYKQVKLVQVHBEKK 501
DB 1343 KPRVKENKVPRLKEHGIELSSPRHSDNPSEGEVKKDDGLEKSPMKKKQK-KKENKENK 1401
QY 502 MELLNMYADK-----KOKLETSKSVEA-----AVIRITCSRTSTQVGD--- 539
DB 1402 EKQSSRKDKEGDKERKSKDKKPKSGDASKSSKRSQGPVHITAGSEVPVIGEED 1461
QY 540 -----LKLDDH-----NYERKPD-----ETKSEKNECLKS-- 564
DB 1462 DDLQDETFSICKERMVRPVKALKQLDKPKGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQMHEVAKKLAEDACWINRIKSWAAKLKVCV 598
DB 1522 DQEHKILWRRNLWIFVSKFTFEDARKLHLKLYMAHKRSQSEEE-----EQKKKDDV 1572
QY 599 PIQSGNNKHF-----SGSSN---ISON 617
DB 1573 ---TGGKKPFRPEASGSSRDLISQS 1595
RESULT 6
AAE36106
ID AAE36106 standard; protein; 1739 AA.
XX AC AAE36106;
XX DT 26-JUN-2003 (first entry)
XX DE Human chromodomain helicase DNA binding protein (CHD) #4.
XX KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
XX KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
XX KW cell proliferative disorder; chromatin organisation modifier domain;
XX KW cytostatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 281..327 /note="Chromodomain"
XX FT Domain 375..433 /note="Chromodomain"
XX FT Domain 487..768 /note="SNF2N domain"
XX FT Domain 831..905 /note="Conserved C-terminal domain"
XX PN WO200298899-A2.
XX PD 12-DEC-2002.
XX 03-JUN-2002; 2002WO-US017466.
XX 05-JUN-2001; 2001US-0296076P.
XX PR 10-OCT-2001; 2001US-0328605P.
XX PR 22-OCT-2001; 2001US-0338733P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX

PA (EXEL-) EXELIXIS INC.
XX
PI Friednan L, Flouman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Llobuin MN;
XX
XX WPI; 2003-156840/15.
XX
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.
XX
XX Claim 13; Page 192-199; 278pp; English.
XX
XX The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human chromodomain
CC helicase DNA binding protein
XX
SQ Sequence 1739 AA;
Query Match 5.1%; Score 162.5; DB 6; Length 1739;
Best Local Similarity 18.9%; Pred No. 0.00071;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
QY 2 ADAPILGSLNPSHDKVHEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
DB 879 ADTVIFDSWNPQNDLQAARAHRIQKQVNIYRLVTGTVTEEIIERAKKMWLDHL 938
QY 50 ILARONRQNAVEN-----LNRLTHALLMGCASYLFDKLDHFHSETPDGVSF 100
DB 939 VIQRMDDTGTITLNNSSGRNSNPKNBELTAILFGAEDLFKELE-----GEES 988
QY 101 EQSIMDGVIEFFSILSSKGEENEVKLLLEAKHAQGVYSSDTLFGEDHIKLSDEES 160
DB 989 EPQEMD-----IDEILRLAETRENEVSTSDTELL-SQFKVANPATWDEEEL-----EER 1038
QY 161 PNIFWSKLLGGKPNMWKYPSTDPQNRKRVOYFE-----GSASPKTG 203
DB 1039 PHKDWDEII-----PEQRKQVEEERQKELEIYMLPRISSTKQAQTNDS 1085
QY 204 DGNVAKRKASDDVTPDTPVDPPVDDDERKASGDHMGALSPKVITLQSSCKSSGTDG 263
DB 1086 DSDTESKROAQRSSASETED-----SDDKK-----PKRRGRPRSVRKOLVFG 1130
QY 264 TLQGN-----DAFLGSMGSHISGIPEDML--ASQDWGKIPDESORLHTVLKPKMAKLC 316
DB 1131 FTDAIRFRFKAYKF-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174
QY 317 QVLHLSDACTSMVGNFLEYIENRIYEEPATTPQAFQIALSWIAALLVKQILSHKESLV 376
DB 1175 ELIH--NSCVSAMQEEYEQLENASEGKPGKR-RGPTIKISGV-QVNVKSIHQHEEFE 1230
QY 377 RANSEL-----AFKSRVEVDYIYILSCMKSLFLEHTQGLQDFCGFNTSK 422
DB 1231 MLHKSIPVDPPEKKYCLTCRVKAAHFDVGVGDDSLRLGLIYEHYGV-NWELIKTDPE 1289
QY 423 QSVVSTKLVSNEISGATVREKINTKMRNSSE-----DESCMTXKR 464
DB 1290 -----LKLTKILPVET--DKKPGKQLQTRADYLLKRLKGLKKGAVTGGEEAKLKR 1342
QY 465 -----CSHYSTATRDIEKTISGIKKYYKQVQKLVQEHBEKK 501
DB 1343 KPRVKENKVPRLKEEHGIELSSPRHSDNPSEEGEVKDDGLEKSPMKKKQK--KKENKENK 1401

QY 502 MELLNMYADK-----KOKLETSKVEA-----AVIRITCSRTSTQVGD--- 539
DB 1402 EKQMSRKQKQEGDKERKSKDKKEKPKSGDAKSSSKSKSQGPVHITAGSEPPVIGEDD 1461
QY 540 -----LKLDDH-----NYERKFD-----EIKSEKNECLKS-- 564
DB 1462 DDLQDQETFSICKERMMPVKALKQDPKGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
QY 565 -----LEQWHEVAKKLADEACWINRIKSWAAKLKVCV 598
DB 1522 DQEHIKLWRRNLWIFVSKFTEFDARKLHLYKMAHKRSQEEB-----EQKKDDV 1572
QY 599 PIOSGNNKHF-----SGSSN---ISON 617
DB 1573 ---TGGKKPFRPRASGSSRDLISQS 1595
RESULT 7
ADK60208
ID ADK60208 standard; protein; 1739 AA.
XX
AC ADK60208;
XX
DT 06-MAY-2004 (first entry)
XX
DE Angiogenesis differentially expressed protein GS-P32.
XX
KW vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
KW antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;
KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;
KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
KW ischemia; angina; myocardial infarction; chronic heart disease;
KW cardiac congestion; macular degeneration; osteoporosis.
XX
OS Homo sapiens.
XX
PN FR2836687-A1.
XX
XX 05-SEP-2003.
XX
XX 11-APR-2002; 2002FR-00004546.
XX
XX 04-MAR-2002; 2002FR-00002717.
XX
XX (GENE-) GENE SIGNAL.
XX (ALMA/) AL MAHMOOD S.
XX
XX Colin S, Schneider C, Al Mahmood S;
XX
XX WPI; 2004-013912/02.
XX N-PSDB; ADK60458.
XX
XX Compositions for diagnosing, prognosing and treating angiogenic disorders
PT including tumor vascularization and heart disease, comprise nucleic acid
PT or polypeptide differentially expressed in angiogenesis.
XX
XX Claim 7; SEQ ID NO 84; 424pp; French.
XX
XX The invention relates to a novel pharmaceutical composition active on
CC angiogenesis comprising an endothelial cell nucleic acid whose expression
CC is induced by an angiogenic factor and inhibited by an angiotatic agent
CC or its complement or fragment, a polypeptide sequence encoded by the
CC nucleic acid or its fragment, a molecule capable of inhibiting expression
CC of the nucleic acid or a molecule which binds to the polypeptide
CC sequence. The invention is used to diagnose, prognose or treat an
CC angiogenic disorder in a mammal, particularly a human. The disorder is
CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
CC endometriosis associated with neovascularization, restenosis due to

CC	angioplasty, overproduction of tissue due to cicatrization, a peripheral
CC	vascular disease, hypertension, vascular inflammation, Raynaud disease,
CC	aneurism, arterial stenosis, thrombophlebitis, ischemia, angina,
CC	myocardial infarction, chronic heart disease, cardiac congestion or
CC	macular degeneration due to age or osteoporosis. This sequence
CC	corresponds to a protein encoded by a differentially expressed DNA used
CC	in the composition of the invention.
XX	
50	Sequence 1739 AA;
	Query Match 5.1%; Score 162.5; DB 8; Length 1739;
	Best Local Similarity 18.9%; Pred. No. 0.00071;
	Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
Qy	2 ADAPILFGSSLPNPHSHDKVHEKIKIESCSRTKIFRLYSVCTVEKA-----L 49
Db	879 ADVVIFDSWNPNQDLQAQARAHRIQKKQVNIYRLVTGTVEEIIRAKKKVLDHL 938
Qy	50 ILARQNWRQKAVN-----LNRSLTALLMNGASYLFDKLDHFHSETPDGVSF 1000
Db	939 VIQRMDDTGTIRLENNSGRNSPNFNKBEALTALKFGAEDLFELE-----GEES 988
Qy	101 EQSTMDGVIHFESSILASKGEENEVKLCLLLEAKHAQGYSSDSTLFGEDHIKLSDEES 160
Db	989 EPQEMD-----IDELRLAETRENEVSTATDELL-SQFKVANPATMEDBEEL-----EER 1038
Qy	161 PNIEWSKLLGGKPNPMKYPSTPORNKRQVQYF-----GSEASPTG 203
Db	1039 PHKWDRII-----PEQRKKVEEERQKELEIYMLPIRSTTKAQTNDS 1085
Qy	204 DGGNAKKRKAASDDVTPRVDPPVDDERKAGCKOHMGALSPKVITTLQSSCKSSGTG 263
Db	1086 DSDTESKRQARSSASETED--SDDKK-----PKRRGRPRSRVKDLVEG 1130
Qy	264 TLQGN-----DAFGLYSMGSHISGIPEDML--ASQDWGKIPDSORLHTVLPRMAKLC 316
Db	1131 FTDAEIRRFIKAYKFF-----GLPIERLECLARD-AELVDKSVADL-----KRLG 1174
Qy	317 QVLHLSDACTSMVGNFLEYVIENHRIYEEEPATTFOAFQIALSWAALLVKOILSHKESLV 376
Db	1175 ELIH--NSCVSAMQVEYEQLENASECKGPKR-RGPTIKISGV-QVNVKSIIQHEEPE 1230
Qy	377 RANSEL-----AFKCSRVEVDYIYSILSCMSKSLFLEHTQGLQFPCFTGNSK 422
Db	1231 MLHKSIPVDPEEKKKYCLTCRVKAAHFDVWGVEDDSRLLLGIVEHGYG-NWELIKTDPE 1289
Qy	423 QSVVSTKLINESLSGATVREKINTKSMRNSR-----DEECMTEKR 464
Db	1290 ----LKLTDKILPVET--DKKPGKQLOLRADYLLKLLKGLKGLKKGAVTGGSEAKLKKR 1342
Qy	465 -----CSHYSTATRDIEKTIISGIIKKYKKQVKVLQVEHEKK 501
Db	1343 KPRVKKENKVPRLKEHGIELSPRHDNFSBEGEVDDGLEKSPMKKKQK-KKENKENK 1401
Qy	502 MELLNMYADK-----KQLETSKSVEA-----AVIRITCSRTSTQVGD--- 539
Db	1402 EKQMSRRKDEGDKERKSKDKKEPKSGDAKSSSKSRSGQPVHITAGSEPPVIGDED 1461
Qy	540 -----LKLDDH-----NYERKFD-----EIKSEKNCEUKS-- 564
Db	1462 DLDQETFSICKERMRPVKKALKQLDKPDKGLNVQEOLEHTRNCLLKIGDRIAECLKAYS 1521
Qy	565 -----LEQMHEVAKKLAEDEACWINRIKSWAAKLKVCV 598
Db	1522 DQEHKLWRNRLWTFVSKFTFDARKLUHKLYKWAHKRSQEE-----EOKKKDPV 1572
Qy	599 PIQSGNNKHF-----SGSSN---ISQN 617
Db	1573 ---TGGKKPFRPEASGSRDSLISQS 1595
RESULT 8	
ADL12347	

Db 1039 PHKDWDEII-----PEEQKKVEEERQKELEEIYMLPRIRSTTKAQTND 1085
Qy 204 DGNNAKRRKASDDVTPRVTDPVDDDERKASGDHMGALSPKVTITLQSSCKSGTGD 263
Db 1086 DSDTESKQRAQRSSASESETED---SDDDKK-----PKRRGRPRSVRKDLVEG 1130
Qy 264 TLDGN-----DAFLYSMSGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLPKMAKLC 316
Db 1131 FTDAETRRFTKAYKKF-----GLPLERLECLARD--AELVDKSVADL-----KRLG 1174
Qy 317 QVLHLSDACTSMVGNFLEVYVNIHRIYEPATTFQAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIHL--NSCVSAMQVEEQLKENASEGKPGKR--RGPTIKISGV-QVNVKSIHQHEEFE 1230
Qy 377 RANSEL-----AFKCSRVEVDYIYISLCMKSLFLEHTQGFQDFCGFNSK 422
Db 1231 MLHKSIPVDPPEKKYKCLTCRVAAHFVWGVEDDSDRLLLGIYEHGYG--NWEIKTDP 1289
Qy 423 QSVVSTKLNVNESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464
Db 1290 -----LKLTDKILPVET--DKKPQGLQTRADYLLKLLKRGLEKKGAVTGSEAKLKGR 1342
Qy 465 -----CSHYSTATRDIEKTIISGIIKKYKQVOKLVQEHHEKK 501
Db 1343 KPRVKKENKVPRLKEBHGIELSPRHSNPNSEGEVKKDGLKSPMKKKQK--KKENK 1401
Qy 502 MELLNMYADK-----KOKLETSKVEA-----AVIRTCSTSTQVGD--- 539
Db 1402 EKQMSRKDKEGDKERKSKDKKEPKSGDAKSSSKRSQGPVHITAGSEVPVIGDED 1461
Qy 540 -----LKLHDH-----NYERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQDQETFSICKMRMPVKALXQLDKPGLNVQEQLEHTRNCLLKIGDRIAECLKAYS 1521
Qy 565 -----LEOMHEVAKKLAEDACWINRIKSWAALKVCV 598
Db 1522 DQSHIKLWRRLNMFVSKPTFDPARKLHKLYMAHKRSQEER-----EQKKKDDV 1572
Qy 599 PIQSGNNKHF-----SGSSN---ISQN 617
Db 1573 ---TGKKKPRPEASGSRDSLISQS 1595
RESULT 9
ID ADK60509
XX ADK60509 standard; protein; 1739 AA.
AC ADK60509;
XX ADK60509;
DT 06-MAY-2004 (first entry)
XX Angiogenesis differentially expressed protein GS-P32.
DE vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
XX antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
XX angiogenesis; endothelial cell; diagnosis; tumor vascularization;
XX retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
XX ovary hyperstimulation; psoriasis; endometriosis; restenosis;
XX angioplasty; cicatrization; peripheral vascular disease; hypertension;
XX vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
XX ischemia; angina; myocardial infarction; chronic heart disease;
XX cardiac congestion; macular degeneration; osteoporosis.
XX Homo sapiens.
OS
XX
XX FR2836686-A1.
PN
XX
XX 05-SEP-2003.
PD
XX
XX 04-MAR-2002; 2002FPR-00002717.
PF
XX
XX 04-MAR-2002; 2002FPR-00002717.
PR

XX (GENE-) GENE SIGNAL.
PA (ALMA/) AL MAHMOOD S.
XX
XX Colin S, Schneider C, Al Mahmood S;
XX WPI; 2004-013911/02.
DR N-PSDB; ADK60759.
XX
XX Compositions containing nucleic acid or polypeptide differentially
PT expressed in angiogenesis are useful to diagnose, prognose and treat
PT angiogenic disorders including tumor vascularization and heart disease.
XX
XX Claim 7; SEQ ID NO 84; 405pp; French.
XX
CC The invention relates to a novel pharmaceutical composition active on
CC angiogenesis comprising an endothelial cell nucleic acid whose expression
CC is induced by an angiogenic factor and inhibited by an angiotatic agent
CC or its complement or fragment, a polypeptide sequence encoded by the
CC nucleic acid or its fragment, a molecule capable of inhibiting expression
CC of the nucleic acid or a molecule which binds to the polypeptide
CC sequence. The invention is used to diagnose, prognose or treat an
CC angiogenic disorder in a mammal, particularly a human. The disorder is
CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
CC endometriosis associated with neovascularization, restenosis due to
CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
CC myocardial infarction, chronic heart disease, cardiac congestion or
CC macular degeneration due to age or osteoporosis. This sequence
CC corresponds to a protein encoded by a differentially expressed DNA used
CC in the composition of the invention.
XX
XX Sequence 1739 AA;
Query Match 5.1%; Score 162.5; DB 8; Length 1739;
Best Local Similarity 18.9%; Pred. No. 0.00071;
Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;
Qy 2 ADAPILFGSLNPSHDVHKVEKIKIESCSERTKIFRLYSVCTVEEKA-----L 49
Db 879 ADTVWIFDSWNPQNDLQAQAHRIQKQKNVIRLVTKGTVEETIERAKKKWLDHL 938
Qy 50 ILARQNRQNAVEN-----LNRLTHALLMGASYLPDKLOHFSHSETPDSQVSP 100
Db 939 VIQRMDTORTILENNSGRNSNPNFKBELTALFKGAEDLFKELE-----GEES 988
Qy 101 EQSIMGVITHEFSSILSSKGEENEVKLCILLLEAKHAQGTYSDDSTLFGDHDKLSDEES 160
Db 989 EPQEMD-----IDEILRLAETRENEVSTSTYDELL--SQPKVANFATMEDEEL---EER 1038
Qy 161 PNIFWSKLGGKNPMWKYPSDTPQRNKRQVQYFE-----GSEAPKGTG 203
Db 1039 PHKDWDEII-----PEEQKKVEEERQKELEEIYMLPRIRSTTKAQTND 1085
Qy 204 DGNNAKRRKASDDVTPRVTDPVDDDERKASGDHMGALSPKVTITLQSSCKSGTGD 263
Db 1086 DSDTESKQRAQRSSASESETED---SDDDKK-----PKRRGRPRSVRKDLVEG 1130
Qy 264 TLDGN-----DAFLYSMSGSHISGIPEDML--ASQDWGKIPDESQRRLHTVLPKMAKLC 316
Db 1131 FTDAETRRFTKAYKKF-----GLPLERLECLARD--AELVDKSVADL-----KRLG 1174
Qy 317 QVLHLSDACTSMVGNFLEVYVNIHRIYEPATTFQAFQIALSWIAALLVKQILSHKESLV 376
Db 1175 ELIHL--NSCVSAMQVEEQLKENASEGKPGKR--RGPTIKISGV-QVNVKSIHQHEEFE 1230
Qy 377 RANSEL-----AFKCSRVEVDYIYISLCMKSLFLEHTQGFQDFCGFNSK 422
Db 1231 MLHKSIPVDPPEKKYKCLTCRVAAHFVWGVEDDSDRLLLGIYEHGYG--NWEIKTDP 1289
Qy 423 QSVVSTKLNVNESLSGATVRDEKINTKSMRNSSE-----DEECMTEKR 464

```

Db 1290 -----LKTDLKLPVET--DKKPGKQLQTRADYLLKLLRKGLEKKGAVTGGESAKLKR 1342
Qy 465 -----
Db 1343 KPRVKENKVPRLKEEHGIELSSPRHSDNPSEEGEVKDDGLKSPMKKKQK-KKENK 1401
Qy 502 MELLNMYADK-----KQKLETSKSVEA-----AVRITCSRTSTQVGD--- 539
Db 1402 EKQMSRRKDKGDKERKKKDKKPKSGDAKSSKSKRSQGPVHITAGSEVPVIGEDD 1461
Qy 540 -----LKLDDH-----NVERKFD-----EIKSEKNECLKS-- 564
Db 1462 DDLQETFSICKMRPVKKALKQDKPKDGLNVOEQLEHTRNCLLKIGDRIACLKAYS 1521
Qy 565 -----LEQMEHVAKKKLAEDACWINRIKSWAAKLKVCV 598
Db 1522 DQEHKILWRRNLWTFVSKTEFDKRLKHLKLYMAHKRSQEE-----EQKKDDV 1572
Qy 599 PIQGNKHP-----SGSSN-----ISQN 617
Db 1573 ---TGKPKPRPEASGSRDSLISQS 1595

```

RESULT 10

ID ADP73132 standard; protein; 1739 AA.

AC ADP73132;

DT 12-AUG-2004 (first entry)

DE Angiogenesis inhibitor human protein sequence, GS-P32.

OS Inhibitor; angiogenesis; antiangiogenic; angiogenesis stimulator;
 KW angiogenic disorder; antiangiogenic; angiogenesis stimulator; cytostatic;
 KW dermatological; antiarthritic; antineoplastic; antineoplastic;
 KW vasotrophic; hypotensive; ophthalmological; antipsoriatic; cardiant;
 KW gene therapy; antiangiogenic gene therapy; tumour vascularisation;
 KW retinopathies; rheumatoid arthritis; Crohn's disease; atherosclerosis;
 KW ovarian hyperstimulation; psoriasis; endometriosis; restenosis;
 KW tissue granulation; peripheral vascular disorder; hypertension;
 KW vascular inflammation; Raynaud's disease; aneurism; arterial restenosis;
 KW thrombophlebitis; lymphadenopathy; lymphedema; ischaemia; angina;
 KW myocardial infarction; chronic heart disease; congestive heart disease;
 KW macular degeneration; human.

OS Homo sapiens.

PN FR2843753-A1.

XX 27-FEB-2004.

XX 20-JUN-2003; 2003FR-00007507.

XX 04-MAR-2002; 2002FR-00002717.

XX (GENE/) GENE S.

XX (ALMS/) AL M S.

PI Colin S, Schneider C, Al MS;

XX WPI; 2004-216677/21.

DR N-PSDB; ADP73081.

XX Antisense nucleic molecule useful as inhibitor of angiogenesis in the
 PT treatment of angiogenic disorders, e.g., rheumatoid arthritis,
 PT atherosclerosis and endometriosis.

XX Claim 1; SEQ ID NO 84; 404pp; French.

XX The invention relates to a novel inhibitor of angiogenesis comprising an
 CC active substance chosen from at least one of a nucleic acid molecule, an

CC antisense nucleic acid molecule, a polypeptide or an antibody. The
 CC invention further comprises: an antisense nucleic acid sequence chosen
 CC from any of the sequences provided in the specification; preparation of a
 CC antibody comprising in vivo or in vitro immunisation of an
 CC immunocompetent animal cell, preferably of a vertebrate and most
 CC preferably of a mammal, with at least one of the polypeptide sequences
 CC chosen from a sequence provided in the specification; a mammalian
 CC expression vector comprising at least one antisense sequence chosen from
 CC an antisense nucleic acid provided in the specification; preparation of a
 CC genetically modified cell, that over- or under-expresses a gene
 CC implicated in an angiogenic disorder, comprising inserting the vector
 CC from above into a mammalian cell; a genetically modified cell that over-
 CC expresses or under-expresses at least one gene involved in angiogenesis
 CC by a nucleic acid sequence chosen from any of ADP73049 to ADP73338, as in
 CC these; preparation of a cell line that stably expresses an expression
 CC vector. The angiogenesis inhibitor has the following activities:
 CC antiangiogenic, angiogenesis stimulator, angiogenesis inhibitor,
 CC cytostatic, dermatological, antiarthritic, antineoplastic,
 CC antiinflammatory, vasotrophic, hypotensive, ophthalmological,
 CC antipsoriatic, and cardiant. The novel nucleic acid molecules of the
 CC invention may be used to treat disorders in gene therapy and antisense
 CC gene therapy. The nucleic acid sequences, proteins and antibodies as part
 CC of the therapeutic compositions are useful in treating a disorder of
 CC angiogenesis chosen from: tumour vascularisation, retinopathies,
 CC rheumatoid arthritis, Crohn's disease, atherosclerosis, ovarian
 CC hyperstimulation, psoriasis, endometriosis associated with
 CC neovascularisation, restenosis due to balloon angioplasty, tissue
 CC granulation due to scarification, peripheral vascular disorders,
 CC hypertension, vascular inflammation, Raynaud's disease, aneurism,
 CC arterial restenosis, thrombophlebitis, lymphadenopathy, lymphedema,
 CC ischaemia, angina, myocardial infarction, chronic heart disease,
 CC congestive heart disease, macular degeneration linked to age and
 CC osteoporosis. This sequence represents a human protein for the creation
 CC of an angiogenesis inhibitor of the invention.

XX SQ Sequence 1739 AA;

Query Match 5.1%; Score 162.5; DB 8; Length 1739;

Best Local Similarity 18.9%; Pred. No. 0.00071;

Matches 152; Conservative 116; Mismatches 259; Indels 279; Gaps 35;

Qy 2 ADAPILPGSLNSHDVHKVEKIKIESCSERTKIFLYSVCTVVEKA-----L 49

Db 879 ADTVIFDSQNPQNDLQAARAHRIQKQVNIYRLVTKGTVEEIERAKKQWLDHL 938

Qy 50 ILARQNMQRKAVEN-----LNESLTHALMWCASLYFDKLDHFSSETPDSGVSF 100

Db 939 VIQRMOTTTGRTILENNSGRNSNPFNKEELTALUKFGAEDLFKELE-----GEES 988

Qy 101 EQSIMGVVIHEFSILSKGGEENEVKLCLLLEAKHAQGYTSSDTLFGEDHIKLSDEES 160

Db 989 EPQEMD-----IDELLALAEETRENEVSTSATDELL-SQFKVANFATMEDEEL---EER 1038

Qy 161 PNTFWSKLGGKPMWKYPSDTPQNRKRVQYVE-----GSASPKTG 203

Db 1039 PHKDWDEII-----PEEQRKVEEERQKELEEIVMLPRISRTTKAQTND 1085

Qy 204 DGGNAKKRKKASDDVTPDPRVTPDPPVDDDERKASKGKHGALSPKVITLQSSCKSSGTDG 263

Db 1086 DSOTESKRQQRSSASETED---SDDDKK-----PRRGRPRSVRKDLVEG 1130

Qy 264 TLDGN-----DAFLGYSMGSHISGIPEDML--ASQDMGKIPDESORLHTVLKPKMAKLC 316

Db 1131 FTDAEIRRFKAYKXF-----GLPLERLECLARD-AELVDKSVADL-----KRLG 1174

Qy 317 QVHLSDACTSMVGNFLEYVNIENRIYEEBPAITFQAFQIALSWTAALLVQILSHKESLV 376

Db 1175 ELIH--NSCVSAMQVEEQLKENASEGKPGKR-RGPTIKISGV-QVNVAKSIHQHEBEFE 1230

Qy 377 RANSEL-----AFKCSRVEVDYIYILSCMKSLFLEHTQGLQDFCGFN 422

Db 1231 MLHKSIPVDEEKKKYCLTCTRVKAAHFDVWGVGDDSRLLLLGIYEHGYG-NWELIKTDPE 1289

RESULT 12
AAU14615
ID AAU14615 standard; protein; 954 AA.
XX AC AAU14615;
XX DT 24-OCT-2001 (first entry)
XX DE Novel bone marrow polypeptide #14.
XX KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
XX KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
XX KW wound healing; nutritional supplement; immune disorder;
XX KW severe combined immunodeficiency; SCID.
XX OS Homo sapiens.
XX PN WO200157187-A2.
XX PD 09-AUG-2001.
XX PP 05-FEB-2001; 2001WO-US003782.
XX PR 03-FEB-2000; 2000US-00496914.
XX PR 20-JUN-2000; 2000US-00598075.
XX PR 19-JUL-2000; 2000US-00620325.
XX PR 30-NOV-2000; 2000US-0250683P.
XX PA (HYSE-) HYSEQ INC.
XX PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
XX PI Ren F, Dmanac RT;
XX DR WPI; 2001-488875/53.
XX DR N-PSDB; AAS22920.
XX PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
XX PT gene therapy.
XX PS Claim 10; Page 215-217; 392pp; English.
XX CC AAU14602-AAU14794 represent novel bone marrow polypeptides of the
XX CC invention. The proteins and corresponding coding sequences may be used in
XX CC the prevention, diagnosis and treatment of diseases associated with
XX CC inappropriate bone marrow polypeptide expression. For example, to treat
XX CC disorders associated with decreased expression by rectifying mutations or
XX CC deletions in a patient's genome that affect the activity of the
XX CC polypeptides by expressing inactive proteins or to supplement the
XX CC patient's own production of the polypeptide. Additionally, the nucleic
XX CC acids may be used to produce the polypeptides, by inserting the nucleic
XX CC acids into a host cell and culturing the cell to express the protein. The
XX CC nucleic acid and its complementary sequences may also be used as DNA
XX CC probes in diagnostic assays to detect and quantitate the presence of
XX CC similar nucleic acid sequences in samples, and therefore which patients
XX CC may be in need of restorative therapy. The proteins may also be used as
XX CC antigens in the production of antibodies against bone marrow proteins and
XX CC in assays to identify modulators of their expression and activity. The
XX CC anti-bone marrow protein antibodies and antagonists may also be used to
XX CC down regulate expression and activity. The antibodies may also be used as
XX CC diagnostic agents for detecting the presence of the protein in samples
XX CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be
XX CC used to regulate haematopoiesis activity, and consequently in the
XX CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
XX CC such as wound healing; as a nutritional supplement; and in treatment of
XX CC immune disorders such as severe combined immunodeficiency (SCID)
XX CC Sequence 954 AA;
XX SQ
Query Match 4.9%; Score 156; DB 4; Length 954;
Best Local Similarity 18.9%; Pred. NO. 0.001;
Matches 131; Conservative 105; Mismatches 233; Indels 224; Gaps 28;
93 TPDSGVFEQIMDGVTHFE8-----SILSSKGGGENEVKLCILLEAKHAQGTSSDSTLF 148
Qy

Db 19 SPGASSLSAVGKG--RQSGGAPGVPEKEENKNEPKICRNWVAPKADPT----- 68
Qy 149 GEDHILKLSDESPNIFWSKILGG-----KNPMWKY-----PSDTPQ-----RNR 187
Db 69 GSEPAKASEKEAPE--DTVDAGEGSRREAEKPKKASALEEGSDASQELEISEHM 125
Qy 188 KRVQYFEGSEASPKTGGGNNAKPKKASDDVTDPRVTDTP-----PVDDDERK 234
Db 126 KEPQLSDSIASDPKSFHGLDFGFRSRISEHLDDVLSVPLGGACROAQOQPLGIEDKDD 185
Qy 235 ASGDHMKALLES-----PKVITLQSSCKS-----STDCDGLD----- 266
Db 186 QSSQDELQSKQSKGLEERYHRLSPPLPHEERAQPPRSLATEEPPQPEQPEWKEAEE 245
Qy 267 -GNDAFGLYSM-----GSHISGIPEDMLA----- 289
Db 246 LGEDSAASLSLQLSLQREQAPSPAACEKGEQHSQAEEELGPGQEEAEDPEEKVAVSTP 305
Qy 290 -----SQDMGKI PDESQRRLHTVLKPKM-----AKLCQ-- 317
Db 306 PVSPEVRSTPEVPAPPEQLSEAAALKAMBEAQAQVLEQDQRHLLSKQEKWQOLREKLCQEE 365
Qy 318 ---VLHL---SDACTSVGNFLEVYIENHRIYEEPPATFOAFQIALSWIALLVKQILSH 371
Db 366 EEEILRLHQKEQSLSLRLRQKAE-----EEEARMEESORLSWLRA-QVOSSTQA 419
Qy 372 KESLVRANSELAFKCSRVEVDYIYSILSCMK---SLFLEHTQGL-----OPDFCGTNSK 422
Db 420 DEDQIRAEQASLQKLREELE-----SQKAERASLEQKNRMLEQKKEIEA-SEKSE 472
Qy 423 QSVVST-----KLVNESLSGATVDEKINTKSMRNSSEDECEMTEKRCRSHYSTATRDI- 475
Db 473 QAALNAAKEKALQOLREQLGE--RKEAVATLEKHSAAE-----LERLCSLEAKHREVV 525
Qy 476 ---EKTISGIKKYYKQVQKLVQEHKEK-----KMLLNMYADKKKLE----- 516
Db 526 SSLQKKIQEAAQKQEEAQLQKCLGQVHRVHQKSYHVAGYEHLSLREKKEQVEGEHER 585
Qy 517 -----TSKSEAAVIRITCSRTSTQVGDLLKLLDHNRYKPFDEIKSEKNEC 561
Db 586 RLDQMKSEHQVMAKAREQYEAERKQRAELLGHLTGELERLQRAHERLETVRQEQHKR 645
Qy 562 LKSLSEQMHVAKKLADEACWINRIKSWAAL 594
Db 646 LEDLRRRHREQERKLDLELDLETRAKDVKARL 678

RESULT 13
ABR61599
ID ABR61599 standard; protein; 2228 AA.
XX AC ABR61599;
XX DT 15-JAN-2004 (first entry)
XX DE Human golgin-245 splice variant 1 polypeptide.
XX KW Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
XX KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
XX OS Homo sapiens.
XX PN WO2003087403-A2.
XX PD 23-OCT-2003.
XX PF 16-APR-2003; 2003WO-EP003958.
XX PR 16-APR-2002; 2002EP-00008553.
XX PR 16-APR-2002; 2002US-0372424P.
XX PA (EVOT-) EVOTEC NEUROSCIENCES GMBH.

XX Von Der Kammer H, Pohlner J;
XX WPI; 2003-845345/78.
DR N-PSDB; ACF58150.
XX
XX Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
XX
XX Claim 18; Fig 5; Opp; English.
XX
XX The invention relates to diagnosing or prognosticating a
CC neurodegenerative disease in a subject, or determining whether a subject
CC is at increased risk of developing the disease. The method involves
CC determining a level and/or an activity of a transcription or translation
CC product of a gene coding for golgin-245 and/or a fragment, derivative or
CC variant in a sample from the subject and comparing the level and/or the
CC activity to a reference value. The composition and methods are useful in
CC diagnosing, preventing and/or treating neurodegenerative diseases, such
CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
CC useful for screening, testing and validating compounds, agents and
CC modulators in the development of diagnostics and therapeutics to treat
CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
CC antibodies are used for detecting the pathological state of a cell in a
CC sample from a subject. The present sequence represents a human golgin-245
CC splice variant 1 polypeptide
XX
XX Sequence 2228 AA;
SQ
Query Match 4.8%; Score 155.5; DB 7; Length 2228;
Best Local Similarity 19.7%; Pred. No. 0.0042;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;
QY 13 NPISHDKVHVEKI-KIESCSB-----RTKIFRLYSVCTVEEKALILARONMKNRAVEN 64
DB 1552 NQOKDIEHKLVLQKLFQBLGEGKONRVK-----EAEKILLLENQVISMKALET 1603
QY 65 LNRLSLTHALMWGA-----SYLFDKLDHPHSETPDGVSFQSDIMDGVITHEFSSILSSK 119
DB 1604 KKLEHVNLSVSKSBEELKALDRLESASAELKRAEKIA-AIKQLLSQMEK 1662
QY 120 GGEENEVKLLLEAKHAQGYSDSTLFGEDHKLSD-FESPNI FWSKLLGGKNPMKY 178
DB 1663 -----BEQYKKGTEHSEL-----NTKQEREREVHILEEKLKSVESQ--- 1702
QY 179 PSPT---PQRNRKRVQFEGSESPKTDGDNKRRKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPSAKNVAAVTEQEDSQ-----GCQVTEEEKISVLQRLTBEKLLQRVGQ 1757
QY 230 DDERKASGKDHMGALSPKVTITLQSSCKSGTDTGLDG-----NDAFGLYSMGSHI 280
DB 1758 EKEETVSHFEMRCYQERLILKEHABAKQHEQDSMIGLHQLEEKKNKYSLI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKIPESQRLHTVLKPKWAKLCQVHLIS-----DACTSNVGNFLE 334
DB 1817 EKEGGKNIOAKONLVNFDVQKTLQ-----EKELTQCIQLEQKIKELDSC-----LV 1864
QY 335 YVTENHRI-YEEPATTFQAFQIALSWI-----AALLVKOILSHKES 374
DB 1865 RQKEVHVEBELTSYKVKLQ-ALQQMDGRNKPTELLEENTEESKSHLVQPKLLSNWEA 1923
QY 375 LVANSLEAPKCSRVEVD-----YIYSILSCMKSLFLEHTQGLQPCFGTNSKQSVST 428
DB 1924 --QHNDLEFKLAGAREKQKLGKIEIVRLQKDLRLMRKEHQEQL-----ILKK 1969
QY 429 KLVNESLGSATVRDEKNTK-----SMRNSSEDECEMTEKCRSHYSTATRIEKTISGKK 484
DB 1970 EYDQE-----REEKIQEQEDLELKHNSLTKQLMREFN-TQLAQKEQELEMTIKETIN 2021
QY 485 KYKQVQKLVQEHKEEKQWELLNMYADKKQKLETSKSVAAVIRITCSTRTSTQVGDKL-- 542

DB 2022 KQAEVEAELLESQEBETNQLLAKAEKDDDLKRTAKRYBEILLDAREEEMTKAVRDLQTL 2081
QY 543 --LDHNYERKFDIEIKSEKNECLKSLEQMEHVAKK-----KLADEACWINKRIKSWAAK 593
DB 2082 EELQKKYQOQLEQENFGNDVNTIMELOTQLAQKTTLISDSKLKEQE--FREQIHLEDR 2139
QY 594 LK-----VCVPIQSGNNKH 607
DB 2140 LKKYKKNVYATTVGTPYKGNLYH 2163
RESULT 14
ABU07445
ID ABU07445 standard; protein; 2230 AA.
XX
AC ABU07445;
XX
DT 28-JAN-2003 (first entry)
XX
DE Protein differentially regulated in prostate cancer #48.
XX
KW Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
XX
OS Homo sapiens.
XX
PN WO200281638-A2.
XX
PD 17-OCT-2002.
XX
PF 08-APR-2002; 2002WO-US010824.
XX
PR 06-APR-2001; 2001US-0281731P.
PR 06-APR-2001; 2001US-0281732P.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Jay G;
XX
DR WPI; 2003-058520/05.
DR N-PSDB; ABX10347.
XX
PT Novel genes which are differentially regulated in prostate cancer, useful
PT for diagnosing prostate cancer in prostate tissue sample and assessing
PT therapeutic or preventive intervention in prostate cancer patients.
XX
PS Claim 1; Page 299-306; 416pp; English.
XX
XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC determining whether the test agent modulates the biological activity of the polypeptide, and
CC is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type

CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (1) can be used as target for therapy or drug
CC discovery. (1) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (1) is useful
CC in therapeutic applications to treat prostate cancer. The identification
CC of specific genes, and groups of genes, expressed in pathways
CC physiologically relevant to prostate cancer permits the definition of
CC functional and disease pathways and the delineation of targets in these
CC pathways which are useful in diagnostic, therapeutic, and clinical
CC applications. This is the amino acid sequence of a protein differentially
CC regulated in prostate cancer
XX
SQ Sequence 2230 AA;

Query Match 4.8%; Score 155.5; DB 6; Length 2230;
Best Local Similarity 19.7%; Pred. No. 0.0042;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;
QY 13 NPSPDHVKEKI-KIESCSE-----RTKIFRLYSVCTVEEKALILARQNRQKAVEN 64
DB 1552 NQKDIHEKLVQKLQHFQELGEEKDNRVK-----EAEKILTLENQVYSMKAELET 1603
QY 65 LNRSLTALLMWGA-----SYLFDKLDHFSSETPDGSGVSPFQSDMGVIEFSSILSSK 119
DB 1604 KKELEHVNLSVKSKEELKALEDRLESESAKLAELKRAEQKIA-AIKQLLSQMEEK 1662
QY 120 GGEENEVKLCLLLEAKHAQGYSSDSTLFGEDHIKLSL-D-EESPNIFWSKLGGRNPMWKY 178
DB 1663 -----EEQYKKGTEHLSL-----NTKLQEREREVHILEBKLSVSSQ--- 1702
QY 179 PSDT---PQRNRKRVQYFEGSEASPTKGDGNAKRRKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPRSAKNAVAAYTEQBEADSQ-----GCVKTYEEKISVLQRLNTEKEKLQVRGQ 1757
QY 230 DDERKASGKHMGALSPKVTLOSSCKSGTGDGLD-----NDAPGLYSMGSHI 280
DB 1758 EKEETVSHPEMRCQYQERLIKLEHAEAKQHEQDSMIGLQEELEEKYKYSLI-VAQHV 1816
QY 281 --SGIPEDMLASQDWGKIPDSQRRLHTVLPKPKMAKLCQVLHLS-----DACTSMVGNFLE 334
DB 1817 EKEGKNKIQAQNLNVFDDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
QY 335 YVIENHRI-YEPPATTFQAFQIALSWI-----AALLVKQILSHKES 374
DB 1865 RQKEVHRVMEBELTSKYEKQLQ-ALQMDGRNKPTELLEENTEESKSHLVQPKLLSNMBA 1923
QY 375 LVANSALAFKCSRVVD-----YVYSILSCMSLFLHTQGLQFCFGTNSQSVST 428
DB 1924 ---QHNDLEFKLAGNERKQKLGKIEVLQKDLRLMRKEHQELE-----ILKK 1969
QY 429 KLVNESLGSATVRDEKINTK-----SMRNSSEDEBCTEKRCSHYSTATRDIEKTISGIKK 484
DB 1970 EYDQE-----REEKIQEQEDLELKHNSLTKQLMREFN-TQLAQKEQLEMTIKETIN 2021
QY 485 KYKQVQKLVQEHKEKQKMLNMVADKKQKLETSKSEVAIVIRTCSTRTSTQVQDLKI-- 542
DB 2022 KAQVEAEELSHOSETNQLLKIAEKDDDLKRTAKRYEBILDAEESMTAKVRDLQTLQ 2081
QY 543 --LQHTYERKFEITSEKNECKLSLEQMHVAKK-----KLAEDACMINRISKWAAK 593
DB 2082 BELQKKYQKLEQEPGNNDVNTMELQTLAQKTLTISDSKLKQE--FREQIHNLDR 2139
QY 594 LK-----VCVPIQSNNKH 607
DB 2140 LKKEKNVYATVGTTPYKGNLYH 2163

RESULT 15
ID ABR61600
XX ABR61600 standard; protein; 2230 AA.
AC ABR61600;
XX

DT 15-JAN-2004 (first entry)
XX Human golgin-245 splice variant 2 polypeptide.
DE
XX Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
KW
XX Homo sapiens.
OS
XX WO2003087403-A2.
PN
XX 23-OCT-2003.
PD
XX 16-APR-2003; 2003WO-EP003958.
PF
XX 16-APR-2002; 2002EP-00008553.
PR
XX 16-APR-2002; 2002US-0372424P.
XX
XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.
PA
XX Von Der Kammer H, Pohlner J;
PI
XX WPI; 2003-845345/78.
DR
XX N-PSDB; ACF58151.
XX
PT Diagnosing or prognosticating a neurodegenerative disease comprises
PT comparing a level or activity of a transcription or translation product
PT of Golgin-245 gene in a test sample with a reference value of a known
PT disease or health status.
XX
PS Claim 18; Fig 7; Opp; English.
XX
CC The invention relates to diagnosing or prognosticating a
CC neurodegenerative disease in a subject, or determining whether a subject
CC is at increased risk of developing the disease. The method involves
CC determining a level and/or an activity of a transcription or translation
CC product of a gene coding for golgin-245 and/or a fragment, derivative or
CC variant in a sample from the subject and comparing the level and/or the
CC activity to a reference value. The composition and methods are useful in
CC diagnosing, preventing and/or treating neurodegenerative diseases, such
CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
CC useful for screening, testing and validating compounds, agents and
CC modulators in the development of diagnostics and therapeutics to treat
CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
CC antibodies are used for detecting the pathological state of a cell in a
CC sample from a subject. The present sequence represents a human golgin-245
CC splice variant 2 polypeptide
XX
SQ Sequence 2230 AA;

Query Match 4.8%; Score 155.5; DB 7; Length 2230;
Best Local Similarity 19.7%; Pred. No. 0.0042;
Matches 135; Conservative 121; Mismatches 267; Indels 161; Gaps 31;
QY 13 NPSPDHVKEKI-KIESCSE-----RTKIFRLYSVCTVEEKALILARQNRQKAVEN 64
DB 1552 NQKDIHEKLVQKLQHFQELGEEKDNRVK-----EAEKILTLENQVYSMKAELET 1603
QY 65 LNRSLTALLMWGA-----SYLFDKLDHFSSETPDGSGVSPFQSDMGVIEFSSILSSK 119
DB 1604 KKELEHVNLSVKSKEELKALEDRLESESAKLAELKRAEQKIA-AIKQLLSQMEEK 1662
QY 120 GGEENEVKLCLLLEAKHAQGYSSDSTLFGEDHIKLSL-D-EESPNIFWSKLGGRNPMWKY 178
DB 1663 -----EEQYKKGTEHLSL-----NTKLQEREREVHILEBKLSVSSQ--- 1702
QY 179 PSDT---PQRNRKRVQYFEGSEASPTKGDGNAKRRKASDDVTDPRVTPPV-----D 229
DB 1703 -SETLIVPRSAKNAVAAYTEQBEADSQ-----GCVKTYEEKISVLQRLNTEKEKLQVRGQ 1757
QY 230 DDERKASGKHMGALSPKVTLOSSCKSGTGDGLD-----NDAPGLYSMGSHI 280
XX

Db 1758 EKEETVSSHPEMRCQYQERLIKLEHAERAKQHQDQSMIGHLQEELEEKNKKYSLI-VAQHV 1816
Qy 281 --SGIPEDMLASQDWGKIPESQRLHTVLKPKWAKLCQVHLHS-----DACTSMVGNFLE 334
Db 1817 EKEGGKNIIQAKONLENVFDVQKTLQ-----EKELTCQILEQKIKELDSC-----LV 1864
Qy 335 YVIEHRI-YEEBATTFOAFQIALSMI-----AALLVKOILSHKES 374
Db 1865 RQKEVHRVENEELTTSYKELQ-ALQQMDGRNKPTELLEENTEESKSHLVOPKLLSNWEA 1923
Qy 375 LVRANSELAFKCSRVEVD-----YIYSIILSCMKSLFLEHTQGLQDFCGTNSKQSVVST 428
Db 1924 ---QHNDLEFKLAGAREKQKIGKEIVRLQKDLRMLRKEHQEULE-----ILKK 1969
Qy 429 KLVNESLSGATVRDEKINTK-----SMRNSSEDECEWTEKCSHYSTATRDIEKTISGIKK 484
Db 1970 EYDQE-----REEKIQEQEDLELKHNSTLKQLMREFN-TQLAQKEQELEMTIKETIN 2021
Qy 485 KYKKQVQKLVQEHHEKKQWELNNMYADKKQKLETSKSVAAVIRITCSRTSTQVGDCLK-- 542
Db 2022 KAGEVEAELESQOEETNQLKKIAEKDDDLKRTAKRYEILDAREEMTAKVRDLQTL 2081
Qy 543 --LDHNYERKPEIKSEKNCLKSLEQMHEVAKK-----KLAEDBACWINRIKSWAAK 593
Db 2082 EELQKKYQOKLEQEEENPGNDNVTIMELOTQAKTTLISDSKLKEQ--FREQIHNELEDR 2139
Qy 594 LK-----VCVPIQSGNNKH 607
Db 2140 LKKYKKNVYATTGTPTYKGNLYH 2163

Search completed: July 12, 2005, 00:58:25
Job time : 73 secs